

# Approximate Bayesian estimation in large coloured graphical Gaussian models

Qiong Li, Xin Gao, Hélène Massam

Department of Mathematics and Statistics, York University

Keywords: Colored  $G$ -Wishart, distributed estimation, double asymptotics, large deviation, marginal model.

## Abstract

Distributed estimation methods have recently been used to compute the maximum likelihood estimate of the precision matrix for large graphical Gaussian models. Our aim, in this paper, is to give a Bayesian estimate of the precision matrix for large graphical Gaussian models with, additionally, symmetry constraints imposed by an underlying graph which is coloured. We take the sample posterior mean of the precision matrix as our estimate. We study its asymptotic behaviour under the regular asymptotic regime when the number of variables  $p$  is fixed and under the double asymptotic regime when both  $p$  and  $n$  grow to infinity. We show in particular, that when the number of parameters of the local models is uniformly bounded, the standard convergence rate we obtain for the asymptotic consistency, in the Frobenius norm, of our estimate of the precision matrix compares well with the rates in the current literature for the maximum likelihood estimate.

## 1 INTRODUCTION

In this paper, we consider graphical Gaussian models with symmetry constraints. Symmetry restrictions for the multivariate Gaussian distribution have a long history dating back to Wilks [1946] and the reader is referred to Gehrmann & Lauritzen [2012] for a complete list of references. Graphical Gaussian models with symmetry restrictions were first considered by Hylleberg et al. [1993]: the symmetry restrictions in that paper could be described by a group action. Subsequently, Andersen [1995] and Madsen [2000] also considered such models. More recently, Højsgaard & Lauritzen [2008] considered graphical Gaussian models with symmetry constraints not necessarily described by a group action. Rather those symmetries are described by coloured graphs  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with skeleton  $G = (V, E)$  where  $V$  is the set of vertices,  $E$  the set of undirected edges,

$\mathcal{V}$  is the set of colour classes for the vertices and  $\mathcal{E}$  the set of colour classes for the edges. The symmetry is given by the equality of certain entries either in the covariance, the correlation or the precision matrices. Models for the multivariate random variable  $X = (X_i, i \in V)$  Markov with respect to  $G$  and with covariance, precision or correlation matrix following equality constraints given by  $\mathcal{G}$  are called coloured graphical Gaussian models. These models have two main advantages. First they may reflect true or imposed symmetries. For example, variables could represent characteristics of twins (see Frets heads data set, Frets [1921]) and therefore the variance of the corresponding variables can be assumed to be equal. Second, since conditional independences imply that certain entries of the precision matrix are set to zero, these restrictions combined with the symmetry restrictions reduce the number of free parameters and facilitate inference in high-dimensional models. Højsgaard & Lauritzen [2008] developed algorithms to compute the maximum likelihood estimate of the covariance, correlation or precision matrix.

In Massam et al. [2015], the authors considered the coloured graphical Gaussian model with symmetry restrictions on the precision matrix and they did so from a Bayesian perspective. In this paper also, we only consider such models which were called RCON models by Højsgaard & Lauritzen [2008]. A pleasant feature of these models is that, given a sample  $X_1, \dots, X_n$  from the coloured graphical Gaussian model with precision matrix  $K$ , the distribution of the sufficient statistics is a natural exponential family with canonical parameter  $K$  and therefore a convenient prior is the Diaconis & Ylvisaker [1979] prior distribution (henceforth abbreviated DY conjugate prior), which we call the coloured  $G$ -Wishart since it is similar to the  $G$ -Wishart which is the DY conjugate prior for graphical Gaussian models Markov with respect to an undirected graph. Massam et al. [2015] gave a method to sample from this posterior (or prior, of course) distribution in order to estimate  $K$  with the sample mean of the posterior DY conjugate distribution. However, as the dimension of the model increases, the computational times also increase and it is not practically feasible to compute the posterior mean for high-dimensional models.

In order to be able to give a Bayesian estimate of the posterior mean of the precision matrix for high-dimensional models, in this paper, we consider distributed estimation, thus providing a Bayesian alternative to the distributed estimation of  $K$  by maximum likelihood. The idea behind distributed estimation is that the estimation of the parameter  $K$  is parsed out to smaller models from which we can estimate part of the parameter

of the initial global model. The estimates of parts of the global parameter are then combined together to yield an estimate of the global model. More precisely, if we want to estimate the precision matrix  $K$  in a graphical Gaussian model with underlying graph  $G = (V, E)$ , for each  $i \in V$  we consider the set of neighbours  $ne(i)$  of  $i$ . At this point, for each  $i$ , we could consider either a local conditional or marginal model: the conditional model of  $X_i$  given  $X_{ne(i)}$  or the marginal model of  $\mathcal{N}_i = \{i\} \cup ne(i)$ . None of the parameters of the local conditional model are equal to part of  $K$  while, as we will show in Section 2, the precision matrix  $K^i$  of the local  $\mathcal{N}_i$ -marginal model is such that

$$K_{il}^i = K_{il}, l \in \mathcal{N}_i. \quad (1)$$

As we will see also in Section 2, in order to make the local marginal model as a natural exponential family while keeping the property (1), we will consider a “relaxed”  $\mathcal{N}_i$ -marginal model. This method was first developed by Meng et al. [2014] for graphical Gaussian models and we adapt it here to coloured graphical Gaussian models.

Having obtained our Bayesian estimate of  $K$  using local marginal models and local coloured  $G$ -Wishart, we will then study its asymptotic properties. We will do so first under the traditional asymptotic conditions, i.e. when the sample size  $n$  goes to infinity and the number of variables  $p$  is fixed and second under the double asymptotic regime when both  $n$  and  $p$  go to infinity.

The study of the asymptotic properties, for  $p$  fixed, of the Bayesian estimate goes back to Bickel & Yahav [1969] who proved the convergence of the normalized posterior density to the appropriate normal density as well as the consistency and efficiency of the posterior mean. Since then, a lot of research has been devoted to Bayesian asymptotics for  $p$  fixed. One of the most recent and well-known work in that area is Ghosal et al. [1995]. For both  $p$  and  $n$  going to infinity, Ghosal [2000] studied the consistency and asymptotic normality, under certain conditions, of the posterior distribution of the canonical parameter for an exponential family when the dimension of the parameter grows with the sample size. Ghosal [2000] also indicates that under additional conditions, the difference between the normalized posterior mean of the canonical parameter and the normalized sample mean tends to 0 in probability.

We will prove in this paper first that, for  $p$  fixed, our estimate is consistent and asymptotically normally distributed, second that, for both  $p$  and  $n$  going to infinity, under certain boundedness conditions and for  $\frac{p^{13}(\log p)^2}{\sqrt{n}} \rightarrow 0$ , our estimate tends, in Frobenius norm, to the true value of the parameter with probability tending to 1. For  $p$  fixed our arguments are classical arguments adapted to our distributed estimate. Under

the double asymptotic regime, there are three main features to our proofs. For each local model, we follow an argument similar to that given in Ghosal [2000]. We therefore need to verify that our DY conjugate prior and our sampling distribution satisfy the conditions and properties assumed by Ghosal [2000] in his arguments. The second feature is that, in the process of proving that the norm of the difference between our estimate and the true value of the parameter tends to 0, we need to prove that asymptotically, our sampling distribution satisfies the so-called cumulant-boundedness condition. To do so, we use an argument similar to that developed by Gao & Carroll [2015] who, in turn, were inspired by the sharp deviation bounds given by Spokoiny & Zhilova [2013] for  $n$  fixed. Finally, we have to combine the results obtained for each local model to show our result for the estimate of the global parameter.

From the condition  $\frac{p^{13}(\log p)^2}{\sqrt{n}} \rightarrow 0$  mentioned above, it would appear that for  $p$  large,  $n$  would have to be extremely large to achieve asymptotic consistency of distributed Bayesian estimate with high probability. This condition is given under the assumption that the number of parameters in each local model is also allowed to grow with  $p$  and  $n$ . When doing estimation, we are actually given the model and it is then reasonable to assume that the number of parameters in the local models is uniformly bounded. Under that more relaxed assumption, as we shall see in Section 5, we obtain a relative rate of growth of  $p$  and  $n$  such that  $\frac{\log^4 p \log \log p}{\sqrt{n}} \rightarrow 0$  which, as will be shown, compares well to the rate given by Meng et al. [2014] for the distributed estimation of the maximum likelihood estimate in graphical Gaussian models.

In Section 2, we recall definitions and basic properties of coloured graphical models and distributed computing. We briefly recall the scheme for sampling from the posterior coloured  $G$ -Wishart. In Section 3, we study the asymptotic properties of our estimate when  $p$  is fixed. In Section 4, we study the asymptotic properties under the double asymptotic regime. In Section 5, we examine how the results of Section 4 are modified when we assume that the number of parameters in the local marginal models is uniformly bounded. In Section 6, we illustrate the efficacy of our method to obtain the posterior mean of  $K$  using several simulated examples. We demonstrate numerically how our method can scale up to any dimension by looking at coloured graphical Gaussian model governed by large coloured cycles and also by a coloured  $10 \times 10$  grid.

## 2 Preliminaries

### 2.1 Coloured graphical models

Let  $X_1, X_2, \dots, X_n$  be independent and identically distributed  $p$ -dimensional random variables following a multivariate normal distribution  $N_p(0, \Sigma)$  with  $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})^t$ ,  $i = 1, 2, \dots, n$ . Let  $K = \Sigma^{-1}$  be the precision matrix and  $G = (V, E)$  be an undirected graph where  $V = \{1, 2, \dots, p\}$  and  $E$  are the sets of vertices and edges, respectively. For  $X = (X_i, i \in V)$ , we say that the distribution of  $X$  is Markov with respect to  $G$  if  $X_i \perp X_j | X_{V \setminus \{i, j\}}$  is implied by the absence of an edge between  $i$  and  $j$  in the graph  $G$ . Such models for  $X$  are called graphical Gaussian models. Since, as it is well-known, conditional independence of the variables  $X_i$  and  $X_j$  is equivalent to  $K_{ij} = 0$ , if we denote  $P_G$  as the cone of positive definite matrices with zero  $(i, j)$  entry whenever the edge  $(i, j)$  does not belong to  $E$ , then the graphical Gaussian model Markov with respect to  $G$  can be represented as

$$\mathcal{N}_G = \{N(0, \Sigma) | K \in P_G\}. \quad (2)$$

Højsgaard & Lauritzen [2008] introduced the coloured graphical Gaussian models with additional symmetry on  $K$  as follows. Let  $\mathcal{V} = \{V_1, V_2, \dots, V_T\}$  form a partition of  $V$  and  $\mathcal{E} = \{E_1, E_2, \dots, E_S\}$  form a partition of  $E$ . If all the vertices belonging to an element  $V_l, l = 1, \dots, T$ , of  $\mathcal{V}$  have the same colour, we say  $\mathcal{V}$  is a colouring of  $V$ . Similarly if all the edges belonging to an element  $E_t, t = 1, \dots, S$ , of  $\mathcal{E}$  have the same colour, we say that  $\mathcal{E}$  is a colouring of  $E$ . We call  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  a coloured graph. Furthermore, if the model (2) is imposed with the following additional restrictions

- (a) if  $m$  is a vertex class in  $\mathcal{V}$ , then for all  $i \in m$ ,  $K_{ii}$  are equal, and
- (b) if  $s$  is an edge class in  $\mathcal{E}$ , then for all  $(i, j) \in s$ ,  $K_{ij}$  are equal,

then the model is defined as a coloured graphical Gaussian model RCON  $(\mathcal{V}, \mathcal{E})$  and denoted as

$$\mathcal{N}_{\mathcal{G}} = \{N(0, \Sigma) | K \in P_{\mathcal{G}}\}$$

where  $P_{\mathcal{G}}$  is the cone of positive symmetric matrix with zero and colour constraints.

We operate within a Bayesian framework. The prior for  $K$  will be the coloured  $G$ -Wishart with density

$$\pi(K | \delta, D) = \frac{1}{I_G(\delta, D)} |K|^{(\delta-2)/2} \exp\{-\frac{1}{2} \text{tr}(KD)\} \mathbf{1}_{K \in P_{\mathcal{G}}},$$

where  $\delta > 0$  and  $D$ , a symmetric positive definite  $p \times p$  matrix, are the hyper parameters of the prior distribution,  $\mathbf{1}_A$  denotes the indicator function of the set  $A$  and  $I_{\mathcal{G}}(\delta, D)$  is the normalizing constant, namely,

$$I_{\mathcal{G}}(\delta, D) = \int_{P_{\mathcal{G}}} |K|^{(\delta-2)/2} \exp\{-\frac{1}{2}tr(KD)\} dK.$$

In the previous expression,  $tr(\cdot)$  represents the trace and  $|K|$  represents the determinant of a matrix  $K$ .

Massam et al. [2015] proposed a sampling scheme for the coloured  $G$ -Wishart distribution. This sampling method is based on Metropolis-Hastings algorithm and Cholesky decomposition of matrices. To develop this sampling method, the authors make the change of variable from  $K$  to  $\Psi = \Phi Q^{-1}$  where  $D^{-1} = Q^t Q$  and  $K = \Phi^t \Phi$  are the Cholesky decomposition of  $D^{-1}$  and  $K$  respectively with  $Q$  and  $\Phi$  upper triangular matrices with real positive diagonal entries. The superscript  $t$  denotes the transpose. Then the zero and colour constraints on the entries of  $K$  associated with a coloured graph  $\mathcal{G}$  determine the sets of free entries in  $\Psi$ . The proposal distribution is a product of normal distributions and chi-square distributions of the free elements in  $\Psi$ . Finally, we complete the non-free elements in  $\Psi$  as a function of the free elements and obtain the sampled  $K = Q^t(\Psi^t \Psi)Q$ .

## 2.2 Local relaxed marginal model

For a given vertex  $i \in V$ , define the set of immediate neighbors of vertex  $i$  as  $ne(i) = \{j | (i, j) \in E\}$ . For each  $i \in V$ , we consider two types of neighbourhood of  $i$ , the so-called one-hop and two-hop neighbourhood. The one-hop neighbourhood  $N_i = \{i\} \cup ne(i)$  is made up of  $i$  and the vertices directly connected to it. The two-hop neighbourhood  $N_i = \{i\} \cup ne(i) \cup \{k | (k, j) \in E, j \in ne(i)\}$  consists of  $i$ , its neighbours and the neighbours of the neighbours. Without risk of confusion, we let  $N_i$  denote either a one-hop or two-hop neighbourhood. Consider the local marginal model for  $X_{N_i} = \{X_v, v \in N_i\}$  which is abbreviated as  $X^i$ . This is a Gaussian model with precision matrix denoted by  $\mathcal{K}^i$ . Then

$$\mathcal{K}^i = (\Sigma_{N_i, N_i})^{-1} = K_{N_i, N_i} - K_{N_i, V \setminus N_i} [K_{V \setminus N_i, V \setminus N_i}]^{-1} K_{V \setminus N_i, N_i}. \quad (3)$$

Based on the collection of vertices  $N_i$  and its complement set  $V \setminus N_i$ , we partition  $N_i$  further into two subsets. One is the buffer set  $B_i = \{j | j \in N_i \text{ and } ne(j) \cap (V \setminus N_i) \neq \emptyset\}$ , which are the vertices having edges connecting to the complement of  $N_i$  in  $V$ . The other is the protected set  $\mathcal{P}_i = N_i \setminus B_i$ , which are the vertices in  $N_i$  that are not directly

connected to  $V \setminus N_i$ . Since the distribution of  $X$  is Markov with respect to  $G$ , then  $X_{\mathcal{P}_i} \perp X_{V \setminus N_i} | X_{B_i}$  and it follows that

$$K_{\mathcal{P}_i, V \setminus N_i} = 0. \quad (4)$$

Then equation (3) becomes

$$\begin{aligned} & \begin{pmatrix} \mathcal{K}_{\mathcal{P}_i, \mathcal{P}_i}^i & \mathcal{K}_{\mathcal{P}_i, B_i}^i \\ \mathcal{K}_{B_i, \mathcal{P}_i}^i & \mathcal{K}_{B_i, B_i}^i \end{pmatrix} \\ = & \begin{pmatrix} K_{\mathcal{P}_i, \mathcal{P}_i} & K_{\mathcal{P}_i, B_i} \\ K_{B_i, \mathcal{P}_i} & K_{B_i, B_i} \end{pmatrix} - \begin{pmatrix} K_{\mathcal{P}_i, V \setminus N_i} \\ K_{B_i, V \setminus N_i} \end{pmatrix} (K_{V \setminus N_i, V \setminus N_i})^{-1} \begin{pmatrix} K_{V \setminus N_i, \mathcal{P}_i} & K_{V \setminus N_i, B_i} \end{pmatrix} \\ = & \begin{pmatrix} K_{\mathcal{P}_i, \mathcal{P}_i} & K_{\mathcal{P}_i, B_i} \\ K_{B_i, \mathcal{P}_i} & K_{B_i, B_i} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & K_{B_i, V \setminus N_i} (K_{V \setminus N_i, V \setminus N_i})^{-1} K_{V \setminus N_i, B_i} \end{pmatrix} \end{aligned}$$

where the 0's in the matrix above follows from the identity (4). Therefore, we obtain the following relationships

$$\begin{aligned} \mathcal{K}_{\mathcal{P}_i, \mathcal{P}_i}^i &= K_{\mathcal{P}_i, \mathcal{P}_i}, & \mathcal{K}_{\mathcal{P}_i, B_i}^i &= K_{\mathcal{P}_i, B_i} \\ \mathcal{K}_{B_i, B_i}^i &= K_{B_i, B_i} - K_{B_i, V \setminus N_i} (K_{V \setminus N_i, V \setminus N_i})^{-1} K_{V \setminus N_i, B_i}. \end{aligned}$$

This shows that the local parameters of  $\mathcal{K}^i$  indexed by  $(\mathcal{P}_i, \mathcal{P}_i)$  and  $(\mathcal{P}_i, B_i)$  are equal to the corresponding global ones but the same does not hold for those indexed by  $(B_i, B_i)$ . This important observation motivates us to use the  $N_i$ -marginal local models to estimate those parameters which are identical in both local and global models.

We denote by  $\mathcal{G}_i$  the coloured graph with vertex set  $N_i$  and edge set

$$E_i = E \cap \{\{\mathcal{P}_i \times \mathcal{P}_i\} \cup \{\mathcal{P}_i \times B_i\} \cup \{B_i \times \mathcal{P}_i\}\} \cup \{B_i \times B_i\}.$$

In  $\mathcal{G}_i$ , the colours of the vertices in  $N_i \setminus B_i$  are the same as the corresponding ones in  $\mathcal{G}$ . The colours of the edges in  $E_i \setminus \{B_i, B_i\}$  are the same as the corresponding ones in  $\mathcal{G}$ . The colours of the vertices in  $B_i$  and the edges in  $B_i \times B_i$  are arbitrary without constraints. Let  $K^i$  denote the precision matrix of this relaxed local marginal model. We thus keep the important relationships

$$\mathcal{K}_{\mathcal{P}_i, \mathcal{P}_i}^i = K_{\mathcal{P}_i, \mathcal{P}_i}, \quad \mathcal{K}_{\mathcal{P}_i, B_i}^i = K_{\mathcal{P}_i, B_i}$$

and have a local Gaussian model with canonical parameter  $K^i$  on which we can put a local coloured  $G$ -Wishart distribution. In each local model Markov with respect to

$\mathcal{G}_i$ ,  $i \in \{1, 2, \dots, p\}$ , we use the method proposed by Massam et al. [2015] to obtain the Bayesian estimator  $\tilde{K}^i$ , the sample posterior mean of  $K^i$  with prior distribution the coloured  $G$ -Wishart.

Next, we will show how to construct a distributed Bayesian estimate by combining local Bayesian estimates. Let  $\theta = (\theta_{V_1}, \theta_{V_2}, \dots, \theta_{V_T}, \theta_{E_1}, \theta_{E_2}, \dots, \theta_{E_S})^t$  denote the global parameter, that is the “free” entries of  $K$  which represent the vertex class or the edge class, and let  $\theta_0$  be its true value. In each local model  $\mathcal{G}_i$ , we define the local parameter as  $\theta^i = (\theta_1^i, \theta_2^i, \dots, \theta_{S_i}^i)^t$ , the vector of free entries of  $K^i$ , and the corresponding local estimator as  $\tilde{\theta}^i$ . The true value of  $\theta^i$  is denoted by  $\theta_0^i$ . Furthermore, we collapse all the local parameters into one vector

$$\bar{\theta} = ((\tilde{\theta}^1)^t, (\tilde{\theta}^2)^t, \dots, (\tilde{\theta}^p)^t)^t$$

and its true value is denoted as  $\bar{\theta}_0$ . After obtaining the local estimators, a distributed estimate of  $\bar{\theta}$  can be constructed as

$$\tilde{\theta}_{V_k} = g_{V_k}(\bar{\theta}) = \frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j=1}^{S_i} \tilde{\theta}_j^i \mathbf{1}_{\theta_j^i = \theta_{V_k}}, \quad k = 1, 2, \dots, T,$$

and

$$\tilde{\theta}_{E_k} = g_{E_k}(\bar{\theta}) = \frac{1}{2|E_k|} \sum_{i \in G_k} \sum_{j=1}^{S_i} \tilde{\theta}_j^i \mathbf{1}_{\theta_j^i = \theta_{E_k}}, \quad k = 1, 2, \dots, S,$$

where  $G_k = \{i | \exists h \in N_i, (i, h) \in E_k\}$ . Define the global distributed Bayesian estimate

$$\tilde{\theta} = g(\bar{\theta}) = (g_{V_1}(\bar{\theta}), g_{V_2}(\bar{\theta}), \dots, g_{V_T}(\bar{\theta}), g_{E_1}(\bar{\theta}), g_{E_2}(\bar{\theta}), \dots, g_{E_S}(\bar{\theta}))^t.$$

### 3 The Bayesian estimator $\tilde{\theta}$ when $p$ is fixed and $n \rightarrow \infty$

Let  $\xrightarrow{\mathcal{L}}$  and  $\xrightarrow{P}$  denote the convergence in distribution and in probability, respectively. In each local model corresponding to the vertex  $i \in \{1, 2, \dots, p\}$ , let  $L^i(\theta^i)$  and  $l^i(\theta^i)$  denote the likelihood and log likelihood, respectively. The Fisher information is denoted by  $I^i(\theta^i) = E_{\theta^i}[\frac{\partial}{\partial \theta^i} l^i(\theta^i | X^i) [\frac{\partial}{\partial \theta^i} l^i(\theta^i | X^i)]^t]$ . Define a  $S_i$ -dimensional vector  $U_{ij} = \frac{1}{\sqrt{n}} [I^i(\theta_0^i)]^{-1} \frac{\partial l^i(\theta^i | X_j^i)}{\partial \theta^i} \Big|_{\theta^i = \theta_0^i}$  for  $j = 1, \dots, n$  and  $i = 1, \dots, p$ , a  $\sum_{i=1}^p S_i$ -dimensional vector  $U_j = (U_{1j}^t, U_{2j}^t, \dots, U_{pj}^t)^t$  and  $\bar{G} = n \text{Cov}(U_1)$ . For each  $r$ ,  $r = 1, 2, \dots, S_i$ , let  $\delta_r^i$  be the  $S_i \times S_i$  indicator matrix with  $(\delta_r^i)_{hl} = 1$  if  $K_{hl}^i = \theta_r^i$  and 0 otherwise. The following theorem shows that the global estimator has the property of asymptotic normality when the number of variables  $p$  is fixed and the sample size  $n$  goes to infinity.



**Theorem 3.1** Let  $\theta_0$ ,  $\tilde{\theta}$ ,  $\bar{\theta}$  and  $\bar{G}$  be defined above. Then

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, A) \quad \text{as } n \rightarrow \infty$$

where  $A = \frac{\partial g(\bar{\theta})}{\partial \theta^t} \bar{G} (\frac{\partial g(\bar{\theta})}{\partial \theta^t})^t$ .

All proofs of theorems and lemmas necessary to their proof are given in the Appendix. We now establish a result similar to Theorem 3.1 but with the MLE replacing the posterior mean. Based on the same local models, we compute the local MLE  $\hat{\theta}^i$  of  $\theta^i$  and obtain a distributed MLE, which is denoted by  $\hat{\theta}$ .

**Theorem 3.2** Let  $\hat{\theta}$  be the distributed MLE. Then

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, A) \quad \text{as } n \rightarrow \infty$$

where  $A$  is defined as in Theorem 3.1 above.

The distributed MLE is computed by the method of Meng et al. [2014] using the local relaxed marginal models defined above. We thus see that the distributed Bayesian estimator  $\tilde{\theta}$  has the same limiting distribution as the distributed MLE  $\hat{\theta}$ .

## 4 The Bayesian estimator under the double asymptotic regime $p \rightarrow \infty$ and $n \rightarrow \infty$

In this section, we study the consistency of the global estimator  $\tilde{\theta}$  when both  $p$  and  $n$  go to infinity. For a vector  $x = (x_1, x_2, \dots, x_p)$ , let  $\|x\|$  stand for its Euclidean norm  $(\sum_{i=1}^p x_i^2)^{1/2}$ . For a square  $p \times p$  matrix  $A$ , let  $\|A\|$  be its operator norm defined by  $\sup\{\|Ax\| : \|x\| \leq 1\}$ , let  $\|A\|_F$  be its Frobenius norm defined by  $\|A\|_F = (\sum_{j=1}^p \sum_{k=1}^p |a_{jk}|^2)^{1/2}$ , and let  $\lambda(A)$ ,  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  be the eigenvalues, the smallest eigenvalues and largest eigenvalues of  $A$ , respectively. The vector obtained by stacking columnwise the entries of  $A$  is denoted by  $\text{vec}(A)$ . Let  $I_p$  be the identity matrix with  $p$  dimension. In the local model Markov with respect to  $\mathcal{G}_i$  as defined in Section 2.2 above, we write the density of  $X_j^i$ ,  $j = 1, 2, \dots, n$ , as

$$f(X_j^i; K^i) = \frac{|K^i|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \text{tr}(K^i X_j^i (X_j^i)^t)\right\}}{(2\pi)^{\frac{p_i}{2}}} \mathbf{1}_{K^i \in P_{\mathcal{G}_i}}$$

where  $p_i = |N_i|$ . The normalized local coloured  $G$ -Wishart distribution of  $K^i$  is denoted by

$$\pi^i(K^i | \delta^i, D^i) = \frac{1}{I_{\mathcal{G}_i}^i(\delta^i, D^i)} |K^i|^{(\delta^i - 2)/2} \exp\left\{-\frac{1}{2} \text{tr}(K^i D^i)\right\} \mathbf{1}_{K^i \in P_{\mathcal{G}_i}},$$

where  $I_{\mathcal{G}^i}(\delta^i, D^i)$  is the normalizing constant. In order to obtain our results, we will follow an argument similar to that of Ghosal [2000] which gives the asymptotic distribution of the posterior mean when both the dimension  $p$  of the model and the sample size  $n$  go to  $\infty$ . Ghosal [2000] considers a random variable  $X$  with density belonging to the natural exponential family

$$f(x; \theta) \propto \exp[x^t \theta - \psi(\theta)],$$

where  $x$  is the canonical statistic,  $\theta$  is the canonical parameter and  $\psi(\theta)$  is the cumulant generating function. To follow the notations of Ghosal [2000], we define an  $S_i$ -dimensional vector

$$Y_j^i = -\frac{1}{2}(tr(\delta_1^i X_j^i (X_j^i)^t), tr(\delta_2^i X_j^i (X_j^i)^t), \dots, tr(\delta_{S_i}^i X_j^i (X_j^i)^t))^t, \quad (5)$$

where  $\delta_1^i, \delta_2^i, \dots, \delta_{S_i}^i$  are indicator matrices for each colour class,  $j = 1, 2, \dots, n$ . The distribution of  $Y_j^i$  is as follows

$$f(Y_j^i; K^i) \propto \exp \left[ -\frac{1}{2} tr(K^i X_j^i (X_j^i)^t) + \frac{1}{2} \log |K^i| \right] = \exp [(Y_j^i)^t \theta^i - \psi(\theta^i)]$$

where  $\psi(\theta^i) = -\frac{1}{2} \log |K^i|$  is the cumulant generating function. From standard properties of natural exponential families, we have that

$$\mu^i = \psi'(\theta_0^i) \quad \text{and} \quad F^i = \psi''(\theta_0^i) \quad (6)$$

are the mean vector and the covariance matrix of  $Y_j^i$ ,  $j = 1, 2, \dots, n$ , respectively. Let  $J^i$  be a square root of  $F^i$ , i.e.  $J^i (J^i)^t = F^i$ . Let

$$V_j^i = (J^i)^{-1} (Y_j^i - E_{\theta^i}(Y_j^i)) \quad (7)$$

be the standardized version of the canonical statistic. Following Ghosal [2000], for any constant  $c$ ,  $c > 0$ , we define

$$B_{1n}^i(c) = \sup \{ E_{\theta^i} |a^t V_j^i|^3 : a \in \mathbb{R}^{S_i}, \|a\| = 1, \|J^i(\theta^i - \theta_0^i)\|^2 \leq \frac{c S_i}{n} \}$$

and

$$B_{2n}^i(c) = \sup \{ E_{\theta^i} |a^t V_j^i|^4 : a \in \mathbb{R}^{S_i}, \|a\| = 1, \|J^i(\theta^i - \theta_0^i)\|^2 \leq \frac{c S_i}{n} \}.$$

Define also

$$u^i = \sqrt{n} J^i (\theta^i - \theta_0^i),$$

then  $\theta^i = \theta_0^i + n^{-1/2}(J^i)^{-1}u^i$ . Therefore, the likelihood ratio can be written as a function of  $u^i$  in the following form

$$Z_n^i(u^i) = \frac{\prod_{j=1}^n f(Y_j^i; \theta^i)}{\prod_{j=1}^n f(Y_j^i; \theta_0^i)} = \exp\{\sqrt{n}(\bar{Y}^i)^t (J^i)^{-1}u^i - n[\psi(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i) - \psi(\theta_0^i)]\},$$

where  $\bar{Y}^i = \frac{1}{n} \sum_{j=1}^n Y_j^i$ . Furthermore, we denote

$$\Delta_n^i = \sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu^i). \quad (8)$$

The following three conditions will be assumed.

- (1) The orders of  $\log p$  and  $\log n$  are the same, i.e.  $\frac{\log p}{\log n} \rightarrow \zeta > 0$  as  $n \rightarrow \infty$ .
- (2) There exists two constants  $\kappa_1$  and  $\kappa_2$  such that  $0 < \kappa_1 \leq \lambda_{\min}(K_0) < \lambda_{\max}(K_0) \leq \kappa_2 < \infty$ .
- (3) For any  $i \in \{1, 2, \dots, p\}$ , the numbers  $\tau^i$  of the entries  $K_{jk}^i$  in the same colour class is bounded.
- (4) As  $p \rightarrow 0$ , the sample size satisfies the rate  $\frac{p^{13}(\log p)^2}{n^{\frac{1}{2}}} \rightarrow 0$ .

**Remark 4.1** Condition (2) implies  $0 < \frac{1}{\kappa_2} \leq \lambda_{\min}(\Sigma_0) < \lambda_{\max}(\Sigma_0) \leq \frac{1}{\kappa_1} < \infty$ . By the interlacing property of eigenvalues, we have that  $0 < \frac{1}{\kappa_2} \leq \lambda_{\min}((\Sigma_0)_{N_i, N_i}) < \lambda_{\max}((\Sigma_0)_{N_i, N_i}) \leq \frac{1}{\kappa_1} < \infty$  where  $N_i$  is defined as in section 2.2. Therefore,  $0 < \kappa_1 \leq \lambda_{\min}((\Sigma_0)_{N_i, N_i})^{-1} < \lambda_{\max}((\Sigma_0)_{N_i, N_i})^{-1} \leq \kappa_2 < \infty$ . By the definition (3), for any  $i \in \{1, 2, \dots, p\}$ , we have  $0 < \kappa_1 \leq \lambda_{\min}(K_0^i) < \lambda_{\max}(K_0^i) \leq \kappa_2 < \infty$ .

Our aim in this section is to prove that under the Conditions (1)-(4) when both  $p$  and  $n$  go to infinity, the distributed estimator  $\tilde{\theta}$  tends to  $\theta_0$  in Frobenius norm with probability tending to 1. Ghosal [2000] considered the consistency of the posterior mean for the exponential family. The rate of the convergence depends on three expressions which added together yield an upper bound of the overall error  $\|\tilde{\theta} - \theta_0\|$ . In each expression, the only random component is  $\|\Delta_n^i\|$  and  $\|\Delta_n^i\| = O_p(p)$ . However, we have an infinite number of local models. In order to use Bonferroni inequality to bound the overall error probability, we need to know the exact tail probability of  $P(\|\Delta_n^i\| > cp)$ , where  $c$  is a constant. This leads us to establish a new large deviation result for  $\|\Delta_n^i\|$  in Lemma 8.3. We now state the asymptotic consistency of our proposed estimator in Theorem 4.1.

**Theorem 4.1** *Under Conditions (1)-(4), there exists a constant  $c^*$  such that*

$$\|\tilde{\theta} - \theta_0\| \leq c^* \frac{p^{\frac{3}{2}}}{\sqrt{n}}$$

*with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2 \log p + \log p\}$ .*

## 5 The Bayesian estimator under the double asymptotic regime when the dimension of the local models bounded

We saw in Section 4 what the asymptotic behaviour of  $\tilde{\theta}$  is when  $S_i$  is unbounded under the double asymptotic regime. In this Section, we assume that  $S_i$  is bounded and we will see that for  $\tilde{\theta}$  to be close to  $\theta_0$ ,  $n$  must grow as a power of  $\log p$  rather than as a power of  $p$ . Indeed, we assume the following conditions:

(4\*) As  $p \rightarrow \infty$ , the sample size satisfies the rate  $\frac{\log^4 p \log \log p}{\sqrt{n}} \rightarrow 0$ .

(5) The number of parameters in each local model is bounded by a constant  $S^*$ , i.e.  $S_i \leq S^*$ ,  $i \in \{1, 2, \dots, p\}$ .

The main result is Theorem 5.1 below.

**Theorem 5.1** *Under Conditions (1), (2), (4\*) and (5), there exists a constant  $c_1^*$  such that*

$$\|\tilde{\theta} - \theta_0\| \leq c_1^* \frac{p^{\frac{1}{2}} \log p}{\sqrt{n}}$$

*with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p + \log p\}$ .*

For the convenience of the reader, we now point out the main difference between the proofs of Theorems 4.1 and 5.1.

(a) Under Conditions (1), (2) and (5), for any  $i \in \{1, 2, \dots, p\}$ , the quantities  $\log |F^i|$  in Proposition 9.3,  $B_{1n}^i(c)$  and  $B_{2n}^i(c)$ ,  $i = 1, 2, \dots, p$ , in Proposition 9.6 are all uniformly bounded because the number of parameters in each local model is uniformly bounded and the eigenvalues of  $K^i$  are uniformly bounded from above and below.

(b) The equivalent of Lemma 8.2 under our new boundedness condition is Lemma 8.8 where  $\|\gamma^i\| < p$  is replaced by the condition  $\|\gamma^i\| < \log p$ .

(c) The equivalent of Lemma 8.3 is Lemma 8.9, the large deviation result is established for  $\|\Delta_n^i\|^2 > 3a^2(\log p)^2$  rather than  $\|\Delta_n^i\|^2 > 3a^2p^2$ .

(d) When  $S_i$  is unbounded, in Theorem 2.1 of Ghosal [2000], the fact that  $\|\Delta_n^i\|^2 > 3a^2p^2$  with probability  $1 - \varepsilon$  implies  $n\|\hat{\theta}^i - \theta_0^i\|^2 > bp^2$  with the same probability  $1 - \varepsilon$ , where  $\hat{\theta}^i$  is the MLE and  $b$  is a constant. For  $S_i$  bounded, using the new large deviation result  $\|\Delta_n^i\|^2 > 3a^2(\log p)^2$  in (c) above, we have the new result of  $n\|\hat{\theta}^i - \theta_0^i\|^2 > b'(\log p)^2$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}(\log p)^2\}$ , where  $b'$  is a constant (See Lemma 8.10).

(e) As a consequence of our choice  $\|\gamma^i\| < \log p$  in (b) above, the threshold  $M(p) = p^2 \log p$  in the proof of Theorem 4.1 can be replaced by  $M(p) = (\log p)^2(\log \log p)$ .

**Remark 5.1** We note that the error bound  $\frac{p^{\frac{1}{2}} \log p}{\sqrt{n}}$  in Theorem 5.1 is smaller than that  $\frac{p^{\frac{3}{2}}}{\sqrt{n}}$  in Theorem 4.1. Also the rate of growth for the sample size is in terms of powers of  $\log p$  rather than  $p$  as in Section 4.

**Remark 5.2** As in Meng et al. [2014], we assume that the graph structure is known. When  $S_i < S^*$ , the error bound in our case is of the order  $\frac{p \log^2 p}{n}$  which compares well with the order  $\frac{p \log p}{n}$  in Meng et al. [2014]. The sample size requirement  $\frac{\log^4 p \log \log p}{\sqrt{n}} \rightarrow 0$  is slightly more demanding than Meng's condition of  $n > \log p$ .

## 6 Simulations

In order to evaluate the performance of our proposed distributed Bayesian estimate of  $K$ , we conduct a number of experiments using simulated data. For each experiment, we compute the distributed estimator using relaxed local marginal models built on the “one-hop” and on the “two-hop” neighbourhoods of each  $i \in \{1, 2, \dots, p\}$ . We choose the coloured  $G$ -Wishart distribution as the prior with hyperparameters  $D^i = I_{p_i}$  and  $\delta^i = 3$  for all  $i \in \{1, 2, \dots, p\}$ . The corresponding estimators are called the MBE-1hop and MBE-2hop estimates of  $K$  respectively. We consider seven different coloured graphical Gaussian models. The underlying graph of three of those models are cycles of length  $p = 20$  with alternate vertex and edge colours as indicated in Figure 1 (a), (b) and (c). Three other models have the same type of underlying coloured graphs but the cycles are of length  $p = 30$ . The underlying graph of the seventh model is a  $10 \times 10$  grid with colours as shown in Figure 1 (d). For both the cycles and the  $10 \times 10$  grid, black edges or vertices indicate that there is no colour constraint. For the sake of comparison, for models with underlying graphs the cycles of order 20 or 30, we also compute the global Bayesian estimate of  $K$ , denoted GBE, using the method given in Massam et al.

[2015]. Since asymptotically, the posterior mean is expected to be close to the maximum likelihood estimate of  $K$ , for all models, we also compute the global MLE of  $K$ , denoted GMLE.

The values of  $(K_{ij})_{1 \leq i, j \leq p}$  used for the simulation for models with underlying graphs as given in Figure 1 (a), (b) and (c) are given in Table 1. For the  $10 \times 10$  coloured grid-graph of Figure 1 (d), we chose  $K_{i+10(j-1), i+1+10(j-1)} = 1$  for  $i = 1, 2, \dots, 9$  and  $j = 1, 2, \dots, 10$ ,  $K_{i+10(j-1), i+10j} = 1 + 0.01i + 0.1j$  for  $i = 1, 2, \dots, 10$  and  $j = 1, 2, \dots, 9$  and  $K_{i,i} = 10 + 0.01i$  for  $i = 1, 2, \dots, 100$ . The posterior mean estimates are based on 5000 iterations after the first 1000 burn-in iterations.

The posterior mean estimates are based on 5000 iterations after the first 1000 burn-in iterations. Table 2 shows the normalized mean square error  $NMSE(K, \hat{K}) = \frac{\|\hat{K} - K\|^2}{\|K\|^2}$  for the six models with the coloured cycles as underlying graphs. Values are averaged over 100 data sets from the normal  $N(0, K^{-1})$  distribution. Standard deviations are shown in parentheses. From these results, we see that our MBE-1hop and MBE-2hop estimates perform very well compared to the global estimate GBE. In Figure 2 we give the graphs of  $NMSE(\hat{K}, K)$  in function of sample size, for different sample sizes ranging from 50 to 100 for the four models with underlying graphs the coloured cycles of length  $p = 20$  and the  $10 \times 10$  grid. We see that the MLE and the GBE consistently yields the smallest and largest NMSE respectively with the MBE-1hop and MBE-2hop in between with the NMSE of the MBE-2hop estimate always smaller than that of the MBE-1hop. As expected, as  $n$  increases, all NMSE tend to the same value.

Computations are performed on a 2 core 4 thread processor with i5-4200U, 2.3 GHZ chips and 8 GB of RAM, running on Windows 8. The average computing times for the estimates of  $K$  are given in minutes for the six models with cycles as underlying graphs. We can see that the computation times for the MBE-1hop and MBE 2-hop are much smaller than for the GBE.

## 7 Conclusion

In this paper, we have given a distributed method to compute the posterior mean of the precision matrix of a coloured graphical Gaussian model, using the DY conjugate prior. To our knowledge, this is the first time that a distributed method has been used in a Bayesian framework. It is also the first time, we believe, that a Bayesian estimate of  $K$  for a high-dimensional coloured graphical Gaussian model, has been given.

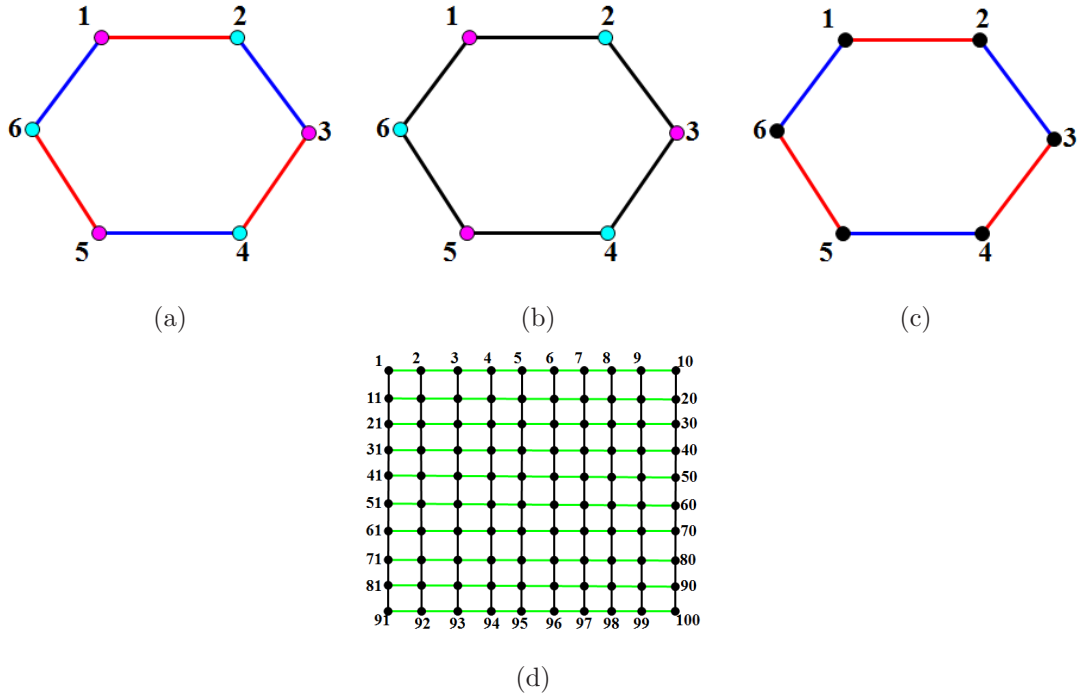
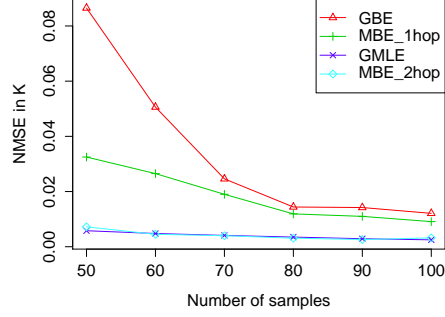
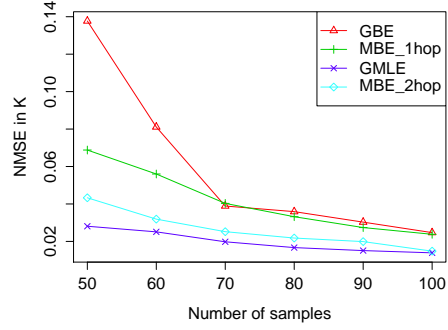


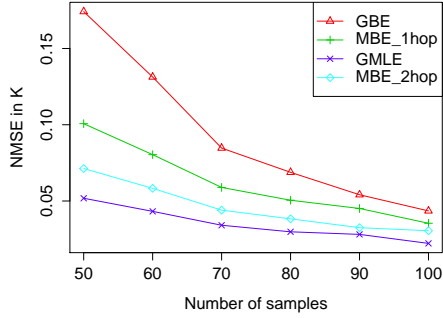
Figure 1: Cycles of length 6 with the three different patterns of colouring that we use for the cycles of length  $p = 20$  and  $p = 30$ . Black vertices or edges indicate different arbitrary colours.



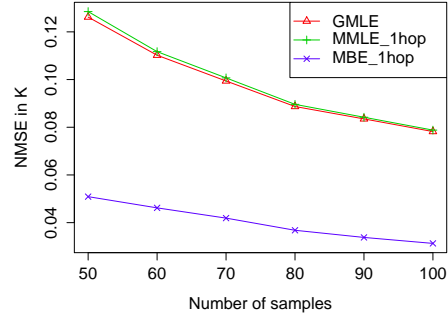
(a)



(b)



(c)



(d)

Figure 2: NMSE in  $K$  for different colored graphical models. (a) NMSE for the colored graph in Figure 1 (a) when  $p = 20$ . (b) NMSE for the colored graph in Figure 1 (b) when  $p = 20$ . (c) NMSE for the colored graph in Figure 1 (c) when  $p = 20$ . (d) NMSE for the colored lattice graph in Figure 1 (d) when  $p = 100$ .



Table 1: The parameters chosen for the matrix  $K$  for producing Figure 1.

parameters	Figure 1 (a)	Figure 1 (b)	Figure 1 (c)
$K_{ii} \ (i = 1, 3, \dots, 2p - 1)$	0.1	0.1	$0.1+0.1i$
$K_{ii} \ (i = 2, 4, \dots, p)$	0.03	0.3	$0.03+0.01i$
$K_{i,i+1} = K_{i+1,i} \ (i = 1, 3, \dots, 2p - 1)$	0.01	$0.01+0.001i$	0.01
$K_{i,i+1} = K_{i+1,i} \ (i = 2, 4, \dots, p - 2)$	0.02	$0.01+0.002i$	0.02
$K_{1p} = K_{p1}$	0.02	0.01	0.02

Table 2:  $NMSE(K, \hat{K})$  for the three coloured models when  $p = 20$  and  $p = 30$ .

NMSE				
$p$	$\mathcal{G}$	MBE_1hop	MBE_2hop	GBE
20	(a)	0.0162 (0.0155)	0.0032 (0.0027)	0.0110 (0.0102)
	(b)	0.0256 (0.0153)	0.0148 (0.0058)	0.0237 (0.0189)
	(c)	0.0375 (0.0283)	0.0305 (0.0142)	0.0308 (0.0241)
30	(a)	0.0098 (0.0070)	0.0017(0.0014)	0.0317 (0.0571)
	(b)	0.0234 (0.0088)	0.0151(0.0054)	0.0482 (0.0533)
	(c)	0.0379 (0.0127)	0.0308 (0.0086)	0.0823 (0.0257)

Table 3: Timing for the three coloured models when  $p = 20$  and  $p = 30$ .

Timing				
$p$	$\mathcal{G}$	MBE_1hop	MBE_2hop	GBE
20	(a)	0.365	3.410	21.875
	(b)	1.047	3.353	16.249
	(c)	0.944	3.054	15.513
30	(a)	1.442	4.952	83.965
	(b)	1.538	4.557	80.255
	(c)	1.504	4.509	79.918

We have first shown that, for our distributed method, the local models should be marginal, rather than conditional, because from these local marginal models, we can directly extract all the components of the precision matrix. Second, we have studied the asymptotic properties of our Bayesian estimate  $\tilde{\theta}$  both when the number  $p$  of variables is fixed and when it is allowed to grow to infinity. When  $p$  is fixed, in our first main result, Theorem 3.1, we prove the asymptotic normality of  $\sqrt{n}(\tilde{\theta} - \theta_0)$  where  $\theta_0$  is the true value of the parameter, using classical methods of asymptotic theory. When both  $p$  and  $n$  tend to infinity, part of our methodology is to extend the method of Ghosal [2000] to our distributed method but we also developed novel methods to prove that the Frobenius norm of  $\tilde{\theta} - \theta_0$  becomes arbitrarily small with probability tending to 1, as  $n$  and  $p$  tend to infinity. Indeed, in Lemma 8.2, parallel to Lemma 1 of Gao & Carroll [2015], we show that if the third derivatives of the cumulant generating functions of  $\Delta_n^i = \sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu_i)$ ,  $i = 1, \dots, p$ , are uniformly bounded, then the subnormality condition (17) given in Lemma 8.2 holds. This, in turn allows us to prove the large deviation result given in Lemma 8.3, which, after being combined for all local models, yields our second main result, Theorem 4.1. Though the convergence rate given in Theorem 4.1 seems to indicate that for this distributed estimate to achieve convergence to  $\theta_0$ ,  $n$  has to be very large, in fact, we show in Theorem 5.1, our third main result, that when we assume that the number of parameters in the local models is uniformly bounded, then the relative growth rate for  $p$  and  $n$  compares well with that given by Meng et al. [2014] under similar conditions. The examples in Section 6, show, through computations, that our method is efficient with sample sizes actually smaller than suggested by our asymptotic results.

## References

- Andersen, H. H. (1995). *Linear and graphical models: for the multivariate complex normal distribution* (Vol. 101). Springer Science and Business Media.
- Bickel, P. J., & Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(4), 257-276.
- Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of statistics*, 7(2), 269-281.

- Frets, G. P. (1921). Heredity of headform in man. *Genetica*, 3(3), 193-400.
- Faraut, J., & Korányi, A. (1994). *Analysis on symmetric cones*. Oxford Mathematical Monographs, Clarendon Press, Oxford.
- Gao, X., & Carroll, R. J. (2015). Data Integration with High Dimensionality. submitted.
- Gehrmann, H., & Lauritzen, S. L. (2012). Estimation of means in graphical Gaussian models with symmetries. *The Annals of Statistics*, 40(2), 1061-1073.
- Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *Journal of Multivariate Analysis*, 74(1), 49-68.
- Ghosal, S., Ghosh, J. K., & Samanta, T. (1995). On convergence of posterior distributions. *The Annals of Statistics*, 2145-2152.
- Højsgaard, S., & Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 1005-1027.
- Hylleberg, B., Jensen, M. B., & Ørnbøl, E. (1993). *Graphical symmetry models*. Aalborg University, Institute for Electronic Systems, Department of Mathematics and Computer Science.
- Lehmann, E., & Casella, G. (1998). *Theory of Point Estimation*. Springer, Berlin.
- Madsen, J. (2000). Invariant normal models with recursive graphical Markov structure. *Annals of statistics*, 1150-1178.
- Massam, H., Li, Q., & Gao, X. (2015). Bayesian precision matrix estimation for graphical Gaussian models with edge and vertex symmetries. arXiv preprint arXiv:1506.04347v1.
- Meng, Z., Wei, D., Wiesel, A., & Hero, A. O. (2014). Marginal likelihoods for distributed parameter estimation of Gaussian graphical models. *Signal Processing, IEEE Transactions on*, 62(20), 5425-5438.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley and Sons.

- Ortega, J. M., & Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables* (Vol. 30). Siam.
- Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16(1), 356-366.
- Spokoiny, V., & Zhilova, M. (2013). Sharp deviation bounds for quadratic forms. *Mathematical Methods of Statistics*, 22(2), 100-113.
- Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 257-281.

## 8 Appendix

In the following, we provide the proofs to all theorems.

**Proof of Theorem 3.1** For any  $i \in \{1, 2, \dots, p\}$ , we have that  $\sqrt{n}(\tilde{\theta}^i - \theta_0^i) = \sqrt{n}(\tilde{\theta}^i - T^i) + \sqrt{n}(T^i - \theta_0^i)$  where  $T^i = \theta_0^i + \frac{1}{n}[I^i(\theta_0^i)]^{-1} \frac{\partial l^i(\theta^i)}{\partial \theta^i} \Big|_{\theta^i = \theta_0^i}$ . It then follows from Theorem 8.3 in Lehmann & Casella [1998] that  $\sqrt{n}(\tilde{\theta}^i - T^i) \xrightarrow{P} 0$ . Furthermore, we have

$$\sqrt{n}(T^i - \theta_0^i) = \frac{1}{\sqrt{n}}[I^i(\theta_0^i)]^{-1} \frac{\partial l^i(\theta^i)}{\partial \theta^i} \Big|_{\theta^i = \theta_0^i} = \sum_{j=1}^n U_{ij}$$

with  $E[U_j] = 0$  for  $j = 1, 2, \dots, n$ . Next, we compute the covariance matrix  $Cov(U_1)$  with  $(i, k)$  entry

$$Cov(U_{i1}, U_{k1}) = \frac{1}{n}[I^i(\theta_0^i)]^{-1} E \left[ \frac{\partial l^i(\theta^i | X_1^i)}{\partial \theta^i} \Big|_{\theta^i = \theta_0^i} \left( \frac{\partial l^k(\theta^k | X_1^k)}{\partial \theta^k} \Big|_{\theta^k = \theta_0^k} \right)^t \right] [I^k(\theta_0^k)]^{-1} \quad (9)$$

Based on the definition of the indicator matrix  $\delta_r^i$ ,

$$\theta_0^i = \left( \frac{1}{|\tau_1^i|} tr(\delta_1^i K_0^i), \frac{1}{|\tau_2^i|} tr(\delta_2^i K_0^i), \dots, \frac{1}{|\tau_{S_i}^i|} tr(\delta_{S_i}^i K_0^i) \right)^t,$$

where  $\tau_r^i$  is the numbers of elements belonging to the  $r$ -th colour class in  $K_0^i$ . Since  $X^i$  has a multivariate normal distribution  $N(0, (K_0^i)^{-1})$ , we have

$$\frac{\partial l^i(\theta^i | X_1^i)}{\partial \theta_j^i} \Big|_{\theta^i = \theta_0^i} = \frac{1}{2} tr(\delta_j^i (K_0^i)^{-1}) - \frac{1}{2} tr(\delta_j^i X_1^i (X_1^i)^t).$$

Therefore, the  $(q, m)$  entry of  $E\left[\frac{\partial l^i(\theta^i|X_1^i)}{\partial \theta^i}\bigg|_{\theta^i=\theta_0^i}(\frac{\partial l^k(\theta^k|X_1^k)}{\partial \theta^k}\bigg|_{\theta^k=\theta_0^k})^t\right]$  in (9) is

$$\begin{aligned} & E\left[\frac{\partial l^i(\theta^i|X_1^i)}{\partial \theta_q^i}\bigg|_{\theta^i=\theta_0^i}\frac{\partial l^k(\theta^k|X_1^k)}{\partial \theta_m^k}\bigg|_{\theta^k=\theta_0^k}\right] \\ &= \frac{1}{4}tr(\delta_q^i \Sigma_0^i) \times tr(\delta_m^k \Sigma_0^k) - \frac{1}{4}tr(\delta_q^i \Sigma_0^i) \times tr(\delta_m^k E[X_1^k (X_1^k)^t]) \\ &\quad - \frac{1}{4}tr(\delta_m^k \Sigma_0^k) \times tr(\delta_q^i E[X_1^i (X_1^i)^t]) + \frac{1}{4}E[tr(\delta_q^i X_1^i (X_1^i)^t) \times tr(\delta_m^k X_1^k (X_1^k)^t)] \end{aligned}$$

where  $\Sigma_0^i = (K_0^i)^{-1}$  and  $\Sigma_0^k = (K_0^k)^{-1}$ . According to Isserlis' Theorem, we have that

$$E[X_{1a}X_{1b}X_{1c}X_{1d}] = (\Sigma_0)_{ab}(\Sigma_0)_{cd} + (\Sigma_0)_{ac}(\Sigma_0)_{bd} + (\Sigma_0)_{ad}(\Sigma_0)_{bc}.$$

Therefore, each entry of  $nCov(U_1)$  is well-defined. By Multivariate Central Limit Theorem, we have  $\sqrt{n}(\bar{\theta} - \bar{\theta}_0) \xrightarrow{\mathcal{L}} N(0, \bar{G})$  as  $n \rightarrow \infty$ , where  $\bar{G} = nCov(U_1)$ . As  $\tilde{\theta} = g(\bar{\theta})$ , based on Delta method, we have that  $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, A)$  where  $A = \frac{\partial g(\bar{\theta})}{\partial \bar{\theta}^t} \bar{G} (\frac{\partial g(\bar{\theta})}{\partial \bar{\theta}^t})^t$ .

■

**Proof of Theorem 3.2** For any  $i \in \{1, 2, \dots, p\}$ , we use the well known result for MLE as follows

$$\sqrt{n}(\hat{\theta}^i - \theta_0^i) = \frac{1}{\sqrt{n}}[I^i(\theta_0^i)]^{-1} \sum_{j=1}^n \frac{\partial l^i(\theta^i|X_j^i)}{\partial \theta^i}\bigg|_{\theta^i=\theta_0^i} + R^i \quad (10)$$

where  $R^i \xrightarrow{p} 0$  as  $n \rightarrow \infty$ . Comparing identity (10) with (9) in Theorem 3.1, the result of Theorem 3.2 follows. ■

**Proof of Theorem 4.1** In this theorem, we study the consistency of  $\tilde{\theta}$  in the context of Frobenius norm. In order to do this, first, we evaluate the norm  $\|\tilde{\theta}^i - \theta_0^i\|^2$  in each local model. Since  $\|\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)\|^2 = n(\tilde{\theta}^i - \theta_0^i)^t (J^i)^t J^i (\tilde{\theta}^i - \theta_0^i) \geq n\lambda_{\min}(F^i)\|\tilde{\theta}^i - \theta_0^i\|^2$ , we obtain

$$\begin{aligned} \|\tilde{\theta}^i - \theta_0^i\|^2 &\leq \frac{1}{n\lambda_{\min}(F^i)} \|\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i)\|^2 \\ &= \frac{1}{n\lambda_{\min}(F^i)} \|\Delta_n^i + \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i\|^2 \quad \text{by Lemma 8.7} \\ &\leq \frac{1}{n\lambda_{\min}(F^i)} \left( \|\Delta_n^i\|^2 + \left\| \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i \right\|^2 \right) \quad (11) \end{aligned}$$

where  $\phi(\cdot; v, \Sigma)$  stands for the multivariate normal density of  $N(v, \Sigma)$  and  $\pi_*^i(u^i)$  stands for the posterior distribution of  $u^i$ . Next, for every element of the vector  $\int u^i [\pi_*^i(u^i) -$

$\phi(u^i; \Delta_n^i, I_{S_i})]du^i$  in (11), we will find out its upper bound. Denote  $u^i = (u_1^i, u_2^i, \dots, u_{S_i}^i)^t$ . Then for the  $j$ -th element of  $\int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$ , we have that

$$\int u_j^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i \leq \int \|u^i\|[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i. \quad (12)$$

Let  $\tilde{Z}_n^i(u^i) = \exp[(u^i)^t \Delta_n^i - \frac{1}{2}\|u^i\|^2]$  and  $M(p) = p^2 \log p$ . According to the argument of Theorem 2.3 in Ghosal [2000], the integral  $\int \|u^i\|[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$  in (12) can be bounded by a sum of three integrals as follows.

$$\begin{aligned} & \int \|u^i\| \times |\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})|du^i \\ & \leq \frac{\int_{\|u^i\|^2 \leq cM(p)} \|u^i\| \cdot |\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)|du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i} \end{aligned} \quad (13)$$

$$+ [\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)du^i]^{-1} \int_{\|u^i\|^2 > cM(p)} \|u^i\| \cdot Z_n^i(u^i)\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)du^i \quad (14)$$

$$+ \int_{\|u^i\|^2 > cM(p)} \|u^i\| \phi(u^i; \Delta_n^i, I_{S_i})du^i, \quad (15)$$

where  $c$  is defined in Lemma 8.5. By Lemmas 8.4, 8.5 and 8.6,  $\int \|u^i\| \times |\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})|du^i$  can be bounded by

$$A(p, n, c) = c_5(c) \frac{p^{13} \log p}{\sqrt{n}} + \exp[-c_9(c)p^2 \log p] + \frac{2}{\sqrt{2\pi}} p^{-4a^2+4} + \sqrt{3a^2} \frac{2}{\sqrt{2\pi}} p^{-4a^2+3}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Consequently,

$$\int u_j^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i \leq A(p, n, c)$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Since the dimension of  $\int u^i[\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})]du^i$  is  $S_i$ , from the inequality (11) and Lemma 8.3, we get

$$\|\tilde{\theta}^i - \theta_0^i\|^2 \leq \frac{1}{\lambda_{\min}(F^i)} \left( \frac{3a^2 p^2}{n} + \frac{S_i}{n} A(p, n, c) \right)$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Finally, we will estimate the Frobenius norm  $\|\tilde{\theta} - \theta_0\|$  for the distributed estimator  $\tilde{\theta}$  in terms of  $\|\tilde{\theta}^i - \theta_0^i\|$  from the local model.

By Proposition 9.1, for any  $i \in \{1, 2, \dots, p\}$ ,  $\lambda_{\min}(F^i) \geq \frac{1}{\kappa_2^2}$ . Therefore, we have

$$\begin{aligned}
\|\tilde{\theta} - \theta_0\| &\leq \|\bar{\theta} - \bar{\theta}_0\| \leq \left( \sum_{i=1}^p \|\tilde{\theta}^i - \theta_0^i\|^2 \right)^{\frac{1}{2}} && \text{by triangle inequality} \\
&\leq \left\{ \sum_{i=1}^p \left[ \frac{1}{\lambda_{\min}(F^i)} \left( \frac{3a^2 p^2}{n} + \frac{S_i}{n} A(p, n, c) \right) \right] \right\}^{\frac{1}{2}} \\
&\leq \left\{ \kappa_2^2 \left[ \frac{3a^2 p^3}{n} + \frac{p^2(p+1)}{2n} A(p, n, c) \right] \right\}^{\frac{1}{2}}
\end{aligned}$$

with probability greater than  $1 - 10.4p \exp\{-\frac{1}{6}p^2\}$  by the Bonferroni inequality. Furthermore, Condition (4) implies  $A(p, n, c) \rightarrow 0$ . Therefore, there exists a constant  $c^*$  such that

$$\begin{aligned}
\|\tilde{\theta} - \theta_0\| &\leq \left\{ \kappa_2^2 \left[ \frac{3a^2 p^3}{n} + \frac{p^2(p+1)}{2n} o(1) \right] \right\}^{\frac{1}{2}} \\
&\leq c^* \frac{p^{\frac{3}{2}}}{\sqrt{n}}
\end{aligned}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2 + \log p\} \rightarrow 1$ . ■

**Proof of Theorem 5.1** The proof follows the same line as that of Theorem 4.1. Our aim is to find the upper bound for the three terms (13), (14) and (15).

1. A bound for (13): Under the Condition (5), the Lipschitz continuity in Proposition 9.5 becomes  $|\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| \leq M_1 S^* \|\theta^i - \theta_0^i\|$  when  $\|\theta^i - \theta_0^i\| \leq \sqrt{\|(F^i)^{-1}\| c M(p)/n}$ . We choose  $M(p) = \log^2 p (\log \log p)$ , then  $\varphi_n'(c) = O(\frac{\log p (\log \log p)^{\frac{1}{2}}}{\sqrt{n}})$  and  $f^i(\|\Delta_n^i\|, c) = O(\frac{\log^3 p (\log \log p)^{\frac{1}{2}}}{\sqrt{n}})$  in Lemma 8.4. Therefore, following the same proof of Lemma 8.4, we have that there exists a constant  $c_5'(c)$  such that  $R_1(\|\Delta_n^i\|, c) \leq c_5'(c) \frac{\log^4 p \log \log p}{\sqrt{n}}$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ .

2. A bound for (14): According to Lemma 8.10 and following the same proof of Lemma 2.2 of Ghosal [2000], on  $\|u^i\|^2 > cM(p)$ , we have  $Z_n^i(u^i) \leq \exp[-\frac{1}{4}c \log^2 p (\log \log p)]$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . Following the same proof as that of Lemma 8.5, there exists a constant  $c$  and a constant  $c_9'(c)$  such that

$$R_2(\|\Delta_n^i\|, c) \leq \exp[-c_9'(c)M(p)]$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ .

3. A bound for (15): According to Lemma 8.6, for  $M(p) = \log^2 p (\log \log p)$ , we have

$$\int_{\|u^i\|^2 > cM(p)} \|u^i\| \phi(u^i; \Delta_n^i, I_{S_i}) du^i \leq \frac{2S_i^2}{\sqrt{2\pi}} e^{-\frac{4a^2 M(p)}{S_i}} + \frac{2\sqrt{3a^2 S_i} \log p}{\sqrt{2\pi}} e^{-\frac{4a^2 M(p)}{S_i}}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ .

Combining the above results, we have

$$A(p, n, c) = c'_5(c) \frac{\log^4 p \log \log p}{\sqrt{n}} + e^{-c'_5(c)M(p)} + \frac{2S_i^2}{\sqrt{2\pi}} e^{-\frac{4a^2 M(p)}{S_i}} + \frac{2\sqrt{3a^2 S_i} \log p}{\sqrt{2\pi}} e^{-\frac{4a^2 M(p)}{S_i}}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . It follows

$$\begin{aligned} \|\tilde{\theta} - \theta_0\| &\leq \left\{ \kappa_2^2 \left[ \frac{p3a^2 \log^2 p}{n} + \frac{S_i}{n} A(p, n, c) \right] \right\}^{\frac{1}{2}} \\ &\leq c_1^* \frac{\sqrt{p} \log p}{\sqrt{n}} \end{aligned}$$

with probability greater than  $1 - 10.4p \exp\{-\frac{1}{6} \log^2 p\}$  by the Bonferroni inequality. This completes the proof. ■

Here we provide the lemmas and their proofs. Additional technical lemmas and propositions are provided in a supplementary file. We let

$$\bar{U}_j^i = (J^i)^{-1} (Y_j^i - \mu^i) \quad (16)$$

for  $i = 1, 2, \dots, p$ , and  $j = 1, 2, \dots, n$ . We now want to show the large deviation result for  $\Delta_n^i$ . To do so, we need to show that the cumulant boundedness condition is satisfied by  $\bar{U}_j^i$  (Lemma 8.1). This will allow us to show that  $\Delta_n^i$  satisfy the exponential moment condition (Lemma 8.2). In Lemma 8.3, we obtain the large deviation result for  $\Delta_n^i$ .

**Lemma 8.1** *For any  $i \in \{1, 2, \dots, p\}$ , there exist constants  $\eta$  and  $C_2$  such that under Condition (2) and (3), for  $\|\gamma^i\| \leq \eta$  and for all  $1 \leq k, l, m \leq S_i$ , the absolute value of all the third derivatives of the cumulant generating function  $G_{\bar{U}_j^i}^i(\gamma^i)$  of  $\bar{U}_j^i$  satisfy*

$$\left| \frac{\partial^3 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \right| \leq C_2, \quad j = 1, 2, \dots, n.$$

**Proof** Let  $Y_j^i$  be defined in (5) of Section 4 and  $G_{Y_j^i}^i(\gamma^i) = \log E(e^{(\gamma^i)^t Y_j^i})$  be the cumulant generating function of  $Y_j^i$ . Let  $\gamma^i$  be a  $S_i$ -dimensional vector, by Theorem 3.2.3 in Muirhead [1982], the moment generating function of  $Y_j^i$  is

$$M^i(\gamma^i) = E\{\exp[(\gamma^i)^t Y_j^i]\} = |I_{p_i} + T^i(\gamma^i) \Sigma_0^i|^{-\frac{1}{2}}$$

where  $T^i(\gamma^i)$  is a  $p_i \times p_i$  matrix with  $T_{\alpha\beta}^i = \gamma_k^i$  if  $K_{\alpha\beta}^i = \theta_k^i$ . Therefore, the cumulant generating function  $G_{Y_j^i}^i(\gamma^i)$  of  $Y_j^i$  is given by

$$G_{Y_j^i}^i(\gamma^i) = \log M^i(\gamma^i) = -\frac{1}{2} \log |I_{p_i} + T^i(\gamma^i) \Sigma_0^i|.$$



It is easy to obtain the first, second and third derivative of the cumulant generating function  $G_{Y_j^i}^i(\gamma^i)$ , which can be expressed as

$$\begin{aligned}\frac{\partial G_{Y_j^i}^i(\gamma^i)}{\partial \gamma_k^i} &= -\frac{1}{2} \text{tr} \left( [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} (\delta_k^i \Sigma_0^i) \right), \\ \frac{\partial^2 G_{Y_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i} &= \frac{1}{2} \text{tr} \left( \delta_k^i \Sigma_0^i [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} (\delta_l^i \Sigma_0^i) [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} \right) \text{ and} \\ \frac{\partial^3 G_{Y_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} &= -\frac{1}{2} \text{tr} \left( \delta_k^i \Sigma_0^i [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} (\delta_m^i \Sigma_0^i) (I_{p_i} + T^i(\gamma^i) \Sigma_0^i)^{-1} (\delta_l^i \Sigma_0^i) \right. \\ &\quad \times [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} + \delta_k^i \Sigma_0^i [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} (\delta_l^i \Sigma_0^i) [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} \\ &\quad \left. \times (\delta_m^i \Sigma_0^i) [I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1} \right),\end{aligned}$$

respectively. First, Condition (2) implies  $\lambda_{\max}(\Sigma_0^i) \leq \frac{1}{\kappa_1}$ . By Proposition 9.2, the absolute value of each element of  $\Sigma_0^i$  is bounded by  $\frac{1}{\kappa_1}$ . Next, by  $\sum_{j=1}^p |\lambda_j(A)| \leq \|A\|_F$  and  $\|AB\| \leq \|AB\|_F \leq \|A\|_F \|B\|$  for any two  $p \times p$  symmetric matrix, we have that  $|\lambda_j(T^i(\gamma^i) \Sigma_0^i)| \leq \|T^i(\gamma^i)\|_F \|\Sigma_0^i\| \leq \eta \frac{1}{\kappa_1}$ . It implies  $1 - \eta \frac{1}{\kappa_1} \leq \lambda_j(I_{p_i} + T^i(\gamma^i) \Sigma_0^i) \leq 1 + \eta \frac{1}{\kappa_1}$ . Moreover, according to Lemma 9.3,  $I_{p_i} + T^i(\gamma^i) \Sigma_0^i$  is a positive definite. Therefore, by Proposition 9.2 again, the absolute value of each element of  $[I_{p_i} + T^i(\gamma^i) \Sigma_0^i]^{-1}$  is bounded. Finally, combining the above results and Condition (3), for any  $i \in \{1, 2, \dots, p\}$ , there exists a constant  $C_1$  such that  $|\frac{\partial^3 G_{Y_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i}| \leq C_1$  for any  $k, m, l$ . Since the cumulant generating function of  $\bar{U}_j^i$  is

$$G_{\bar{U}_j^i}^i(\gamma^i) = \log E[e^{(\gamma^i)^t (J^i)^{-1} (Y_j^i - \mu^i)}] = G_{Y_j^i}^i((J^i)^{-1} \gamma^i) - (\gamma^i)^t (J^i)^{-1} \mu^i.$$

It follows that there exists a constant  $C_2$  such that  $|\frac{\partial^3 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i}| \leq C_2$  for  $\|\gamma^i\| \leq \eta$ . ■

**Lemma 8.2** Let  $\Delta_n^i$  and  $\bar{U}_j^i$  be as defined in (8) of Section 4 and (16), respectively. Let  $C_2$  be as in Lemma 8.1. Then, under Condition (2)-(4), for any arbitrary constant  $a$  such that  $a^2 > 1$ , we have that if  $\frac{C_2 p^3}{3\sqrt{n}} \leq a - 1$ , then as  $n \rightarrow \infty$ ,

$$G_{\Delta_n^i}^i(\gamma^i) = \log(E\{\exp[(\gamma^i)^t \Delta_n^i]\}) \leq a^2 \|\gamma^i\|^2 / 2 \quad \text{for} \quad \|\gamma^i\| < p. \quad (17)$$

**Proof** By a Taylor expansion of  $G_{\bar{U}_j^i}^i(\gamma^i)$  around 0, there exists a vector  $\gamma^{i,*}$  on the line segment between 0 and  $\gamma^i$  such that

$$\begin{aligned}G_{\bar{U}_j^i}^i(\gamma^i) &= G_{\bar{U}_j^i}^i(0) + \sum_{k=1}^{S_i} \left( \frac{\partial G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i} \Big|_{\gamma^i=0} \right) \gamma_k^i + \frac{1}{2} \sum_{k=1}^{S_i} \sum_{l=1}^{S_i} \left( \frac{\partial^2 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i} \Big|_{\gamma^i=0} \right) \gamma_k^i \gamma_l^i \\ &\quad + \frac{1}{6} \sum_{k=1}^{S_i} \sum_{l=1}^{S_i} \sum_{m=1}^{S_i} \left( \frac{\partial^3 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \Big|_{\gamma^i=\gamma^{i,*}} \right) \gamma_k^i \gamma_l^i \gamma_m^i.\end{aligned}$$

Since  $\bar{U}_j^i$  has zero mean and identity covariance matrices, then  $\frac{\partial G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i} \big|_{\gamma^i=0} = 0$ ,  $\frac{\partial^2 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i} \big|_{\gamma^i=0} = 1$  for  $k = l$  and  $\frac{\partial^2 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i} \big|_{\gamma^i=0} = 0$  for  $k \neq l$ . Furthermore, since  $G_{\bar{U}_j^i}^i(0) = 0$ , we have

$$G_{\bar{U}_j^i}^i(\gamma^i) = \frac{1}{2}(\gamma^i)^t \gamma^i + \frac{1}{6} \sum_{k=1}^{S_i} \sum_{l=1}^{S_i} \sum_{m=1}^{S_i} \left( \frac{\partial^3 G_{\bar{U}_j^i}^i(\gamma^i)}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \big|_{\gamma^i=\gamma^{i,*}} \right) \gamma_k^i \gamma_l^i \gamma_m^i.$$

By the definition (8) of Section 4, we have  $\Delta_n^i = \frac{1}{\sqrt{n}} \sum_{j=1}^n \bar{U}_j^i$ . Since the moment generating function of  $\bar{U}_j^i$  is  $\exp G_{\bar{U}_j^i}^i(\gamma^i)$ , then the moment generating function of  $\Delta_n^i$  is

$$\begin{aligned} E[e^{(\gamma^i)^t \Delta_n^i}] &= E[e^{(\gamma^i)^t \frac{1}{\sqrt{n}} \sum_{j=1}^n \bar{U}_j^i}] = \prod_{j=1}^n E[e^{(\frac{\gamma^i}{\sqrt{n}})^t \bar{U}_j^i}] \\ &= \exp \left\{ \frac{1}{2}(\gamma^i)^t \gamma^i + \frac{1}{6} \frac{1}{\sqrt{n}} \sum_{k=1}^{S_i} \sum_{l=1}^{S_i} \sum_{m=1}^{S_i} \left( \frac{\partial^3 G_{\bar{U}_j^i}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \big|_{\gamma^i=\gamma^{i,*}} \right) \gamma_k^i \gamma_l^i \gamma_m^i \right\}. \end{aligned}$$

Since  $\|\gamma^{i,*}\| < \|\gamma^i\|$ , we have  $\|\frac{\gamma^{i,*}}{\sqrt{n}}\| < \|\frac{\gamma^i}{\sqrt{n}}\| < \frac{p}{\sqrt{n}}$ . Moreover, Condition (4) implies  $\frac{p}{\sqrt{n}} \rightarrow 0$ , and thus  $\|\frac{\gamma^i}{\sqrt{n}}\| \leq \eta$  for  $n$  large enough. Therefore, by Lemma 8.1, there exists a constant  $C_2$  such that  $\left| \frac{\partial^3 G_{\bar{U}_j^i}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \right| \leq C_2$ . It follows

$$\begin{aligned} E[e^{(\gamma^i)^t \Delta_n^i}] &\leq \exp \left\{ \frac{1}{2}(\gamma^i)^t \gamma^i + \frac{1}{6} \frac{C_2}{\sqrt{n}} \sum_{k=1}^{S_i} \sum_{l=1}^{S_i} \sum_{m=1}^{S_i} \gamma_k^i \gamma_l^i \gamma_m^i \right\} \\ &= \exp \left\{ \frac{1}{2}(\gamma^i)^t \gamma^i \left[ 1 + \frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \right] \right\}. \end{aligned}$$

Therefore, for any arbitrary constant  $a$  such that  $a^2 > 1$ , if  $\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \leq a^2 - 1$ , then we have

$$\log E[e^{(\gamma^i)^t \eta^i}] \leq a^2 \|\gamma^i\|^2 / 2.$$

Actually, the inequality  $\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \leq a^2 - 1$  holds under Condition (4). Since  $\|\gamma^i\| < p$ , we have  $|\gamma_m^i| \leq \|\gamma^i\| < p$  for any  $1 \leq m \leq S_i$ . Therefore, according to Condition (4), we have

$$\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i = O\left(\frac{p^3}{\sqrt{n}}\right) = o(1).$$

It implies  $\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \leq a^2 - 1$  for any constant  $a$  with  $a^2 > 1$ . ■

**Lemma 8.3** *Under Condition (2)-(4), for any  $i \in \{1, 2, \dots, p\}$  and  $n$  sufficiently large, there exists a constant  $a$ ,  $a^2 > 1$ , such that*

$$P\{\|\Delta_n^i\|^2 > 3a^2 p^2\} \leq 10.4 \exp\{-\frac{1}{6}p^2\}$$

where  $\Delta_n^i$  is defined as in (8) of Section 4.

**Proof** According to Lemma 8.2, we have

$$\log(E\{\exp[(\gamma^i)^t \Delta_n^i]\}) \leq a^2 \|\gamma^i\|^2 / 2 \quad \text{for} \quad \|\gamma^i\| \leq p$$

where  $a$  is a constant with  $a^2 > 1$ . Let  $g = ap$  and  $t_1^i = a\gamma^i$ , then the subsequent inequality holds

$$\log(E\{\exp[(t_1^i)^t \frac{\Delta_n^i}{a}]\}) \leq \|t_1^i\|^2 / 2 \quad \text{for} \quad \|t_1^i\| \leq g.$$

Next we apply the large deviation result from Corollary 3.2 in Spokoiny & Zhilova [2013]. Following the notations in Spokoiny & Zhilova [2013], we introduce  $w_c^i$  satisfying the equation  $\frac{w_c^i(1+w_c^i)}{[1+(w_c^i)^2]^{\frac{3}{2}}} = gS_i^{-1/2}$ . Based on  $w_c^i$ , we define  $x_c^i = 0.5S_i[(w_c^i)^2 - \log(1+(w_c^i)^2)]$ . Since  $g^2 = a^2 p^2 > \frac{p^2+p}{2} \geq S_i$ , by the arguments in Spokoiny & Zhilova [2013], we have  $x_c^i > \frac{1}{4}g^2 = \frac{1}{4}a^2 p^2$ . Let  $x = \frac{1}{6}p^2$ , then  $\frac{S_i}{6.6} \leq \frac{p^2+p}{2 \times 6.6} < x < x_c^i$ . By Corollary 3.2 in Spokoiny & Zhilova [2013], the following inequality holds

$$P(\|\frac{\Delta_n^i}{a}\|^2 \geq S_i + 6.6 \times \frac{1}{6}p^2) \leq 2e^{-\frac{1}{6}p^2} + 8.4e^{-x_c^i},$$

which implies  $P(\|\frac{\Delta_n^i}{a}\|^2 \geq 3p^2) \leq 10.4e^{-\frac{1}{6}p^2}$ . Hence,  $P(\|\Delta_n^i\|^2 \geq 3a^2 p^2) \leq 10.4e^{-\frac{1}{6}p^2}$ , which means  $\|\Delta_n^i\|^2 = O_p(p^2)$ . ■

The next four lemmas are used to complete the proof of Theorem 4.1.

**Lemma 8.4** *Under Conditions (2)-(4), for any given  $i \in \{1, 2, \dots, p\}$  and for any given constant  $c$ , there exists a constant  $c_5(c)$  such that*

$$\frac{\int_{\|u^i\|^2 \leq cM(p)} \|u^i\| \cdot |\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)Z_n^i(u^i) - \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i)| du^i}{\int \pi^i(\theta_0^i)\tilde{Z}_n^i(u^i) du^i} \leq c_5 \frac{p^{13} \log p}{\sqrt{n}} \quad (18)$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ .

**Proof** Let  $Q^i$  denote the set  $\{u^i; \|u^i\|^2 \leq cM(p)\}$ . We get that

$$\begin{aligned}
& [\int \pi^i(\theta_0^i) \tilde{Z}_n^i(u^i) du^i]^{-1} \int_{Q^i} \|u^i\| \cdot |\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i) Z_n^i(u^i) - \pi^i(\theta_0^i) \tilde{Z}_n^i(u^i)| du^i \\
&= [\int \pi^i(\theta_0^i) \tilde{Z}_n^i(u^i) du^i]^{-1} \int_{Q^i} \|u^i\| \cdot \left| \frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} Z_n^i(u^i) - \tilde{Z}_n^i(u^i) \right| \pi^i(\theta_0^i) du^i \\
&\leq \frac{\sup_{u^i \in Q^i} \left\{ \|u^i\| \cdot \left| \frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1 \right| \right\} \int_{Q^i} Z_n^i(u^i) du^i}{\int \tilde{Z}_n^i(u^i) du^i} + \frac{\int_{Q^i} \|u^i\| \cdot |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)| du^i}{\int \tilde{Z}_n^i(u^i) du^i}.
\end{aligned}$$

Since  $cM(p) \geq \|u^i\|^2 = \|\sqrt{n}J^i(\theta^i - \theta_0^i)\|^2 \geq n\lambda_{\min}(F^i)\|\theta^i - \theta_0^i\|^2$ , then  $\|\theta^i - \theta_0^i\| \leq \sqrt{\frac{cM(p)\|(F^i)^{-1}\|}{n}}$ . By Proposition 9.1, we have  $\kappa_1^2 \leq \|(F^i)^{-1}\| \leq \kappa_2^2$ . Based on Condition (4),  $\frac{p^2 \log p}{n} \rightarrow 0$ . Therefore,  $\|\theta^i - \theta_0^i\| \rightarrow 0$ . Using the fact  $|e^x - 1| \leq 2|x|$  for sufficiently small  $|x|$  and Proposition 9.5, we obtain

$$\sup_{u^i \in Q^i} \left\{ \|u^i\| \cdot \left| \frac{\pi^i(\theta_0^i + n^{-1/2}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} - 1 \right| \right\} \leq 2\sqrt{cM(p)}M_1p\|\theta^i - \theta_0^i\| \leq \frac{2cM_1\kappa_2M(p)p}{\sqrt{n}}$$

where  $M_1$  is a constant. We also have that

$$\begin{aligned}
\frac{\int_{Q^i} Z_n^i(u^i) du^i}{\int \tilde{Z}_n^i(u^i) du^i} &= \frac{\int_{Q^i} \tilde{Z}_n^i(u^i) du^i + \int_{Q^i} [Z_n^i(u^i) - \tilde{Z}_n^i(u^i)] du^i}{\int \tilde{Z}_n^i(u^i) du^i} \\
&\leq 1 + \left( \int \tilde{Z}_n^i(u^i) du^i \right)^{-1} \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)| du^i.
\end{aligned}$$

According to Lemma 2.3 in Ghosal [2000], we can obtain

$$\left( \int \tilde{Z}_n^i(u^i) du^i \right)^{-1} \int_{Q^i} |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)| du^i \leq f^i(\|\Delta_n^i\|, c) \quad (19)$$

where

$$\begin{aligned}
f^i(\|\Delta_n^i\|, c) &= \varphi_n^i(c)[p_i^2 + (1 - 2\varphi_n^i(c))^{-1}\|\Delta_n^i\|^2] \left(1 - 2\varphi_n^i(c)\right)^{-(\frac{p_i^2}{2}+1)} \\
&\quad \times \exp \left\{ \frac{\varphi_n^i(c)\|\Delta_n^i\|^2}{1 - 2\varphi_n^i(c)} \right\},
\end{aligned}$$

and

$$\varphi_n^i(c) = \frac{1}{6} [n^{-\frac{1}{2}}(cM(p))^{\frac{1}{2}} B_{1n}^i(0) + n^{-1}cM(p)B_{2n}^i(c\frac{M(p)}{S_i})].$$

Furthermore, since  $\|u^i\| \leq \sqrt{cM(p)}$ , by the inequality (19), it is easy to see that

$$\frac{\int_{Q^i} \|u^i\| \cdot |Z_n^i(u^i) - \tilde{Z}_n^i(u^i)| du^i}{\int \tilde{Z}_n^i(u^i) du^i} \leq \sqrt{cM(p)}f(\|\Delta_n^i\|, c).$$

Combining the above results, we can show that the LHS in (18) is bounded by

$$R_1(\|\Delta_n^i\|, c) = \frac{2cM_1\kappa_2M(p)p}{\sqrt{n}}[1 + f^i(\|\Delta_n^i\|, c)] + \sqrt{cM(p)}f^i(\|\Delta_n^i\|, c).$$

According to Proposition 9.6, we have  $B_{1n}^i(0) = O(p^9)$  and  $B_{2n}^i(c\frac{M(p)}{S_i}) = O(p^{12})$ . Therefore, there exist two constants  $c_1$  and  $c_2$  such that

$$\varphi_n^i(c) \leq \frac{1}{6}[\frac{(cM(p))^{\frac{1}{2}}c_1p^9}{\sqrt{n}} + n^{-1}cM(p)c_2p^{12}] = \frac{1}{6}\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}[\sqrt{c}c_1 + c_2c\frac{p^4\sqrt{\log p}}{\sqrt{n}}] \quad (20)$$

Since the first term in (20) is the dominating term, then there exists a constant  $c_3(c)$  such that  $\varphi_n^i(c) \leq c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}$ . By Condition (4), we have that  $\varphi_n^i(c) \rightarrow 0$ . Furthermore, using the fact  $(1-x)^{-1} \leq 2$  and  $-\log(1-x) \leq 2x$  for sufficiently small  $x$ , we have  $[1 - 2\varphi_n^i(c)]^{-1} \leq 2$  and  $e^{-(\frac{p_i^2}{2}+1)\log(1-2\varphi_n^i(c))} \leq e^{(\frac{p_i^2}{2}+1)4\varphi_n^i(c)}$ . Therefore, the following inequality holds

$$f^i(\|\Delta_n^i\|, c) \leq \varphi_n^i(c)[p^2 + 2\|\Delta_n^i\|^2] \exp\{(\frac{p^2}{2} + 1)4\varphi_n^i(c)\} \exp\{2\varphi_n^i(c)\|\Delta_n^i\|^2\}.$$

According to Lemma 8.3, we see that  $P(\|\Delta_n^i\|^2 \leq 3a^2p^2) > 1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Therefore,

$$f^i(\|\Delta_n^i\|, c) \leq c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}[p^2 + 6a^2p^2] \exp\left\{c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(6a^2p^2 + 2p^2 + 4)\right\}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . By Condition (4), we have that  $\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(6a^2p^2 + 2p^2 + 2) \rightarrow 0$ . Therefore,  $\exp\left\{c_3(c)\frac{p^{10}\sqrt{\log p}}{\sqrt{n}}(4a^2p^2 + 2p^2 + 2)\right\} < 2$ . It follows

$$f^i(\|\Delta_n^i\|, c) \leq 2(1 + 6a^2)c_3(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Let  $c_4(c) = 2(1 + 6a^2)c_3(c)$ , then  $f^i(\|\Delta_n^i\|, c) \leq c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Furthermore, we can get

$$\begin{aligned} R_1(\|\Delta_n^i\|, c) &= \frac{2cM_1\kappa_2M(p)p}{\sqrt{n}}[1 + c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}}] + \sqrt{cM(p)}c_4(c)\frac{p^{12}\sqrt{\log p}}{\sqrt{n}} \\ &= \frac{2cM_1\kappa_2p^3 \log p}{\sqrt{n}} + \frac{p^{13} \log p}{\sqrt{n}}[c_4(c)\frac{2cM_1\kappa_2p^2 \log^{\frac{1}{2}} p}{\sqrt{n}} + \sqrt{c}c_4(c)] \quad (21) \end{aligned}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . It is easy to see that the third term in (21) is the dominating term. Therefore, there exists a constant  $c_5(c)$  such that  $R_1(\|\Delta_n^i\|, c) \leq c_5(c) \frac{p^{13} \log p}{\sqrt{n}}$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . The proof is completed. ■

**Lemma 8.5** *Under Condition (2)-(4), there exists a constant  $c$  large enough and a constant  $c_9(c)$  such that for any given  $i \in \{1, 2, \dots, p\}$ ,*

$$\frac{\int_{\|u^i\|^2 > cM(p)} \|u^i\| \pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i) Z_n^i(u^i) du^i}{\int \pi^i(\theta_0^i) \tilde{Z}_n^i(u^i) du^i} \leq \exp[-c_9(c)p^2 \log p]$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ .

**Proof** Let

$$\begin{aligned} R_2(\|\Delta_n^i\|, c) &= \frac{\int_{\|u^i\|^2 > cM(p)} \|u^i\| \pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i) Z_n^i(u^i) du^i}{\int \pi^i(\theta_0^i) \tilde{Z}_n^i(u^i) du^i} \\ &= \frac{\int_{\|u^i\|^2 > cM(p)} \|u^i\| \frac{\pi^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1}u^i)}{\pi^i(\theta_0^i)} Z_n^i(u^i) du^i}{(2\pi)^{S_i/2} \exp[\frac{\|\Delta_n^i\|^2}{2}]}. \end{aligned}$$

According to Lemma 2.2 in Ghosal [2000], we have that  $Z_n^i(u^i) \leq \exp[-\frac{1}{4}cp^2 \log p]$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Let  $\pi_0^i(\theta^i)$  denotes the non-normalized local coloured  $G$ -Wishart distribution. Then we obtain that

$$\begin{aligned} &R_2(\|\Delta_n^i\|, c) \\ &\leq \frac{\exp[-\frac{1}{4}cp^2 \log p]}{(2\pi)^{S_i/2} \exp[\frac{\|\Delta_n^i\|^2}{2}]} \\ &\quad \times \int_{\|\sqrt{n}J^i(\theta^i - \theta_0^i)\|^2 > cM(p)} \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \frac{\pi_0^i(\theta^i)}{\pi_0^i(\theta_0^i)} n^{S_i/2} |J^i| d\theta^i \\ &\leq \exp[\frac{S_i}{2} \log n + \frac{1}{2} \log |F^i| - \frac{1}{4}cp^2 \log p - \log \pi_0^i(\theta_0^i)] \\ &\quad + \log \int_{\|\sqrt{n}J^i(\theta^i - \theta_0^i)\|^2 > cM(p)} \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \end{aligned} \tag{22}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . By Proposition 9.4 and Lemma 9.5,

we have that

$$R_2(\|\Delta_n^i\|, c) \leq \exp\left[\frac{S_i}{2} \log n + \frac{1}{2} \log |F^i| - \frac{1}{4} c p^2 \log p + \frac{1}{2} p_i \kappa_2 - \frac{\delta^i - 2}{2} p_i \log \kappa_1 + M_7 p^2 \log p\right]$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} p^2\}$ . By Condition (1),  $\log n$  and  $\log p$  are of the same order. Furthermore, Proposition 9.3 implies  $\log |F^i| = O(p^2)$ . Therefore, there exists a constant  $c_6$  such that  $\log |F^i| \leq c_6 p^2$ . It follows the RHS in (22) is bounded by the following term

$$\exp\left[\frac{p(p+1)}{4} \log p + \frac{1}{2} c_6 p^2 - \frac{1}{4} c p^2 \log p + \frac{1}{2} p_i \kappa_2 - \frac{\delta^i - 2}{2} p_i \log \kappa_1 + M_7 p^2 \log p\right]$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} p^2\}$ . Furthermore, there exists a constant  $c_8$  such that

$$R_2(\|\Delta_n^i\|, c) \leq \exp\left[\frac{p(p+1)}{4} \log p - \frac{1}{4} c p^2 \log p + M_7 p^2 \log p + c_8 p^2 \log p\right]$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} p^2\}$ . We can choose a constant  $c$  big enough such that  $c_9(c) = \frac{1}{4} - \frac{1}{4} c + c_8 + M_7 < 0$ . It immediately implies  $R_2(\|\Delta_n^i\|, c) \leq \exp[-c_9(c) p^2 \log p]$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} p^2\}$ . ■

**Lemma 8.6** *Under Condition (2)-(4), for any given  $i \in \{1, 2, \dots, p\}$  and for any constant  $c$  such that  $c > 11a^2$  and  $a^2 > 1$ , we have*

$$\int_{\|u^i\|^2 > cM(p)} \|u^i\| \phi(u^i; \Delta_n^i, I_{S_i}) du^i \leq \frac{2}{\sqrt{2\pi}} p^{-4a^2+4} + \sqrt{3a^2} \frac{2}{\sqrt{2\pi}} p^{-4a^2+3}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} p^2\}$ .

**Proof** First we observe that

$$\begin{aligned} & \int_{\|u^i\|^2 > cM(p)} \|u^i\| \phi(u^i; \Delta_n^i, I_{S_i}) du^i \\ & \leq \int_{\|u^i\|^2 > cM(p)} (\|u^i - \Delta_n^i\|) \phi(u^i; \Delta_n^i, I_{S_i}) du^i + \int_{\|u^i\|^2 > cM(p)} \|\Delta_n^i\| \phi(u^i; \Delta_n^i, I_{S_i}) du^i. \end{aligned}$$

Let  $v^i = u^i - \Delta_n^i$ , since  $\|v^i\|^2 + \|\Delta_n^i\|^2 \geq \|v^i + \Delta_n^i\|^2 = \|u^i\|^2 > cM(p)$ , then immediately  $\|v^i\|^2 > cM(p) - \|\Delta_n^i\|^2$ . By Lemma 8.3, we can see that  $\|\Delta_n^i\|^2 \leq 3a^2p^2$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$  with  $a^2 > 1$ . As  $c$  is chosen that  $c > 11a^2$ , we can get  $\|v^i\|^2 > cM(p) - \|\Delta_n^i\|^2 > cM(p) - 3a^2p^2 > (11a^2 - 3a^2)p^2 \log p = 8a^2p^2 \log p$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Thus the following inequality holds with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ .

$$\begin{aligned} & \int_{\|u^i\|^2 > cM(p)} (\|u^i - \Delta_n^i\|) \phi(u^i; \Delta_n^i, I_{S_i}) du^i \\ &= \int_{\|v^i + \Delta_n^i\|^2 > cM(p)} \|v^i\| \phi(v^i; 0, I_{S_i}) dv^i \leq \sum_{j=1}^{S_i} \int_{\|v^i\|^2 > 8a^2M(p)} |v_j^i| \phi(v^i; 0, I_{S_i}) dv^i, \end{aligned}$$

where  $v_j^i$  is the  $j$ -th element of  $v^i$ . We also have that

$$\begin{aligned} & \sum_{j=1}^{S_i} \int_{\|v^i\|^2 > 8a^2M(p)} |v_j^i| \phi(v^i; 0, I_{S_i}) dv^i \\ & \leq \sum_{j=1}^{S_i} \sum_{k=1}^{S_i} \int_{\mathbb{R}^{S_i-1}} \int_{(v_k^i)^2 > 8a^2 \frac{M(p)}{S_i}} |v_j^i| \phi(v^i; 0, I_{S_i}) dv^i \leq 2p^4 \frac{1}{\sqrt{2\pi}} p^{-4a^2} \end{aligned}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ , and

$$\int_{\|u^i\|^2 > cM(p)} \|\Delta_n^i\| \phi(u^i; \Delta_n^i, I_{S_i}) du^i \leq \sqrt{3a^2} p^3 \frac{2}{\sqrt{2\pi}} p^{-4a^2}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6}p^2\}$ . Hence, the desired result follows. ■

**Lemma 8.7** *For a given  $i \in \{1, 2, \dots, p\}$ , we have*

$$\sqrt{n}J^i(\tilde{\theta}^i - \theta_0^i) = \Delta_n^i + \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i$$

where  $\pi_*^i(u^i)$  is the posterior distribution of  $u^i$ .

**Lemma 8.8** *(Parallel to Lemma 8.2) Let  $\Delta_n^i$  and  $\bar{U}_j^i$  be as defined in (8) of Section 4 and (16), respectively. Let  $C_2$  be defined as in Lemma 8.1. Then under Conditions (2), (4\*) and (5), for any arbitrary constant  $a$  such that  $a^2 > 1$ , we have that if  $\frac{C_2 \log p}{3\sqrt{n}} \leq a - 1$ , for  $\|\gamma^i\| < \log p$  and  $n$  sufficiently large,*

$$G_{\Delta_n^i}^i(\gamma^i) = \log(E\{\exp[(\gamma^i)^t \Delta_n^i]\}) \leq a^2 \|\gamma^i\|^2 / 2.$$



**Lemma 8.9** (Parallel to Lemma 8.3) Under Condition (2), (4\*) and (5), for any  $i \in \{1, 2, \dots, p\}$  and  $n$  sufficiently large, there exists a constant  $a$ ,  $a^2 > 1$ , such that

$$P\{\|\Delta_n^i\|^2 > 3a^2 \log^2 p\} \leq 10.4 \exp\{-\frac{1}{6} \log^2 p\}$$

where  $\Delta_n^i$  is defined as in (8) of Section 4.

**Lemma 8.10** Let  $\hat{\theta}^i$  be the MLE of  $\theta^i$  in the  $i$ -th local model. Under Condition (2), (4\*) and (5), for any  $i \in \{1, 2, \dots, p\}$ ,

$$\sqrt{n}\|J^i(\hat{\theta}^i - \theta_0^i)\| \leq c' \log p$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ , where  $c' = 1.2\sqrt{3}a \frac{\lambda_{max}(F^i)}{\lambda_{min}^2(F^i)}$ .

## 9 Supplementary file

In this section we provide the proofs of the propositions and lemmas that we have used in the Appendix.

**Proposition 9.1** Let  $F^i$  be defined in definition (6) of Section 4 for any  $i \in \{1, 2, \dots, p\}$ , then under Condition (2), we have that

$$\frac{1}{\kappa_2^2} \leq \lambda_{min}(F^i) \leq \lambda_{max}(F^i) \leq \frac{1}{\kappa_1^2}.$$

**Proof** Let  $G^i$  be the Fisher information matrix for the uncolored graphical models e.g.  $G^i = \psi_u''(\theta^i)$  where  $\psi_u(\theta^i) = (-\frac{1}{2} \log |K^i| + \frac{p_i}{2} \log(2\pi)) \mathbf{1}_{K^i \in P_{G^i}}$ . Let  $\tau$  and  $\varpi$  be the numbers of eigenvalues of  $G^i$  and  $F^i$ . Since  $F^i$  is a linear projection of  $G^i$  onto the space of uncolored symmetric matrices, then  $\tau > \varpi$ . Under Condition (2) and by Proposition 9.7, for any  $l$ ,  $1 \leq l \leq \varpi$ , we have

$$\frac{1}{\kappa_2^2} \leq \min\{\frac{1}{\lambda_j(G^i)\lambda_k(G^i)} | 1 \leq j, k \leq \tau\} \leq \lambda_l(F^i) \leq \max\{\frac{1}{\lambda_j(G^i)\lambda_k(G^i)} | 1 \leq j, k \leq \tau\} \leq \frac{1}{\kappa_1^2}.$$

■

**Proposition 9.2** For any  $i \in \{1, 2, \dots, p\}$ , let  $K_{\alpha\beta}^{i,0}$  be the  $(\alpha, \beta)$  entry of  $K_0^i$ . Under Condition (2), we have  $|K_{\alpha\beta}^{i,0}| \leq \kappa_2$ .

**Proof** By Condition (2), we have  $\lambda_{max}(K_0^i) \leq \kappa_2$  for any  $i \in \{1, 2, \dots, p\}$ . Therefore,  $\kappa_2 - \lambda_j(K_0^i)$ ,  $j = 1, 2, \dots, p_i$ , are the eigenvalues of  $\kappa_2 I_{p_i} - K_0^i$ . Since  $\lambda_{max}(K_0^i) \leq \kappa_2$ , then  $\kappa_2 \geq \lambda_j(K_0^i)$ ,  $j = 1, 2, \dots, p_i$ . It follows that  $\kappa_2 I_{p_i} - K_0^i$  is a positive semidefinite

matrix. Since the diagonal elements of a positive semidefinite  $\kappa_2 I_{p_i} - K_0^i$  are all non negative, then  $\kappa_2 - K_{\alpha\alpha}^{i,0} \geq 0, \alpha = 1, 2, \dots, p_i$ . It follows  $0 < K_{\alpha\alpha}^{i,0} \leq \kappa_2$ . Since  $K_0^i$  is a positive definite matrix, then each 2 by 2 principal sub matrices

$$\begin{pmatrix} K_{\alpha\alpha}^{i,0} & K_{\alpha\beta}^{i,0} \\ K_{\beta\alpha}^{i,0} & K_{\beta\beta}^{i,0} \end{pmatrix}$$

of  $K_0^i$  are positive definite. Therefore,  $K_{\alpha\alpha}^{i,0} K_{\beta\beta}^{i,0} - (K_{\alpha\beta}^{i,0})^2 > 0$ , from which we get  $|K_{\alpha\beta}^{i,0}| < (K_{\alpha\alpha}^{i,0} K_{\beta\beta}^{i,0})^{1/2} < \kappa_2$ . ■

The next four propositions provide the properties of  $\log |F^i|$ , the colored  $G$ -Wishart prior, the third and fourth moments of the normalized  $Y_j^i$ .

**Proposition 9.3** *Under Condition (3), for any  $i \in \{1, 2, \dots, p\}$ , we have the trace of  $F^i$  satisfies  $\text{tr}(F^i) = O(p^2)$  and the determinant  $|F^i|$  satisfies  $\log |F^i| = O(p^2)$ .*

**Proof** Since  $\frac{\partial^2 \psi(\theta^i)}{\partial \theta_j^i \partial \theta_k^i} = \frac{1}{2} \text{tr}(\delta_j^i \Sigma_0^i \delta_k^i \Sigma_0^i)$ , then  $\text{tr}(F^i) = \frac{1}{2} \sum_{j=1}^{S_i} \text{tr}((\delta_j^i \Sigma_0^i)^2)$ . Furthermore, by Condition (3),  $\text{tr}(\delta_j^i \Sigma_0^i)$  is bounded. Therefore,  $\text{tr}((\delta_j^i \Sigma_0^i)^2)$  is bounded. It follows

$$\text{tr}(F^i) = \frac{1}{2} \sum_{j=1}^{S_i} \text{tr}((\delta_j^i \Sigma_0^i)^2) \leq \frac{1}{2} \frac{p_i(p_i + 1)}{2} \text{tr}((\delta_j^i \Sigma_0^i)^2) \leq \frac{1}{2} \frac{p(p + 1)}{2} \text{tr}((\delta_j^i \Sigma_0^i)^2) = O(p^2).$$

Next, let us consider  $\log |F^i|$ . Since  $|F^i| = \prod_{j=1}^{S_i} \lambda_j(F^i) \leq \left( \frac{\sum_{j=1}^{S_i} \lambda_j(F^i)}{S_i} \right)^{S_i} = \left( \frac{\text{tr}(F^i)}{S_i} \right)^{S_i}$ , then

$$\log |F^i| \leq S_i \log \frac{\text{tr}(F^i)}{S_i} \leq \frac{p_i(p_i + 1)}{2} \log \frac{\frac{1}{2} \frac{p_i(p_i + 1)}{2} \text{tr}((\delta_j^i \Sigma_0^i)^2)}{\frac{p_i(p_i + 1)}{2}} = O(p^2).$$

The proposition is proved. ■

**Proposition 9.4** *Under Condition (2), for any  $i \in \{1, 2, \dots, p\}$ , we have*

$$\log \pi_0^i(K_0^i) \geq -\frac{1}{2} p_i \kappa_2 + \frac{\delta^i - 2}{2} p_i \log \kappa_1$$

when  $D^i = I_{p_i}$ .

**Proof** The non-normalized colored  $G$ -Wishart distribution can be rewritten as

$$\begin{aligned} \pi_0^i(K_0^i) &= \exp \left\{ -\frac{1}{2} \text{tr}(K_0^i I_{p_i}) + \frac{\delta^i - 2}{2} \log |K_0^i| \right\} \\ &= \exp \left\{ -\frac{1}{2} \sum_{j=1}^{p_i} \lambda_j(K_0^i) + \frac{\delta^i - 2}{2} \log \prod_{j=1}^{p_i} \lambda_j(K_0^i) \right\} \\ &\geq \exp \left\{ -\frac{1}{2} p_i \kappa_2 + \frac{\delta^i - 2}{2} p_i \log \kappa_1 \right\}. \end{aligned}$$

The last inequality due to Condition (2). ■

**Proposition 9.5** (*Lipschitz continuity*) For any  $i \in \{1, 2, \dots, p\}$  and any constant  $c$ , there exists a constant  $M_1$  such that

$$|\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| \leq M_1 p \|\theta^i - \theta_0^i\|$$

when  $\|\theta^i - \theta_0^i\| \leq \sqrt{\|(F^i)^{-1}\|cM(p)/n} \rightarrow 0$ .

**Proof** Let  $\pi_0^i(\theta^i)$  be the non-normalized colored  $G$ -Wishart distribution for the local model. By mean value theorem, we have

$$\begin{aligned} |\log \pi^i(\theta^i) - \log \pi^i(\theta_0^i)| &= |\log \pi_0^i(\theta^i) - \log \pi_0^i(\theta_0^i)| = |(\theta^i - \theta_0^i)^t \frac{\partial \log \pi_0^i(\theta^i)}{\partial \theta^i} |_{\theta^i = \check{\theta}^i}| \\ &= \|\theta^i - \theta_0^i\| \cdot \sqrt{\sum_{j=1}^{S_i} \left[ -\frac{1}{2} \text{tr}(\delta_j^i D^i) + \frac{\delta^i - 2}{2} \text{tr}(\delta_j^i (\check{K}^i)^{-1}) \right]^2}, \end{aligned}$$

where  $\check{\theta}^i$  is the point on the line segment joining  $\theta^i$  and  $\theta_0^i$ . Since  $\|\theta^i - \theta_0^i\| \rightarrow 0$ , then  $(\check{K}^i)^{-1} \rightarrow (K_0^i)^{-1}$ . According to Condition (2) and Proposition 9.2, each entry of  $(K_0^i)^{-1}$  is uniformly bounded, then using the similar proof of Lemma 9.1, each entry  $(\check{K}^i)^{-1}$  is uniformly bounded. Therefore, there exists a constant  $M_1$  such that

$$\sqrt{\sum_{j=1}^{S_i} \left[ -\frac{1}{2} \text{tr}(\delta_j^i D^i) + \frac{\delta^i - 2}{2} \text{tr}(\delta_j^i \Sigma_0^i) \right]^2} \leq \sqrt{\frac{p_i(1+p_i)}{2}} M_1^2 \leq \sqrt{\frac{p(1+p)}{2}} M_1^2 = M_1 p.$$

■

**Proposition 9.6** For any  $i \in V$ , let  $Y_j^i$  and  $V_j^i$  be defined in (5) and (7) of Section 4, respectively. Then  $B_{1n}^i(c) = O(p^9)$  and  $B_{2n}^i(c) = O(p^{12})$ .

**Proof** Let  $B_{\alpha\beta}$  be the  $(\alpha, \beta)$  entry of  $(J^i)^{-1}$ . Define  $b = \max\{|B_{\alpha\beta}|; \alpha, \beta \in \{1, 2, \dots, S_i\}\}$ . Then for the vectors  $Y_j^i = (Y_{j1}^i, Y_{j2}^i, \dots, Y_{jS_i}^i)^t$  and  $a = (a_1, a_2, \dots, a_{S_i})^t$ , the following property holds for  $h = 1, 2, 3, 4$

$$E_{\theta^i} |a^t (J^i)^{-1} Y_j^i|^h \leq E_{\theta^i} \left[ (|a_1|, |a_2|, \dots, |a_{S_i}|) \begin{pmatrix} b \sum_{k=1}^{S_i} |Y_{jk}^i| \\ b \sum_{k=1}^{S_i} |Y_{jk}^i| \\ \vdots \\ b \sum_{k=1}^{S_i} |Y_{jk}^i| \end{pmatrix} \right]^h = E_{\theta^i} \left[ \left( b \sum_{k=1}^{S_i} |Y_{jk}^i| \right) \sum_{k=1}^{S_i} |a_k| \right]^h. \quad (23)$$

According to Cauchy-Schwarz inequality, we have that

$$E_{\theta^i} |a^t(J^i)^{-1} Y_j^i|^h \leq E_{\theta^i} \left[ b \left( \sum_{k=1}^{S_i} |Y_{jk_k}^i| \right) \sqrt{S_i} \|a\| \right]^h \leq b^h (S_i)^{h/2} E_{\theta^i} \left[ \sum_{k_1=1}^{S_i} \dots \sum_{k_h=1}^{S_i} |Y_{jk_{k_1}}^i| \dots |Y_{jk_{k_h}}^i| \right] \quad (24)$$

According to Lemma 9.1, each entry of  $\theta^i$  is bounded when  $\|J^i(\theta^i - \theta_0^i)\|^2 \leq \frac{cS_i}{n} \rightarrow 0$ .

By Lemma 9.2, we have  $E_{\theta^i} [|Y_{jk_1}^i| \dots |Y_{jk_h}^i|]$  is bounded for  $h = 1, 2, 3, 4$ . Therefore,  $E_{\theta^i} |a^t(J^i)^{-1} Y_j^i|^h = O(p_i^{3h})$ . Similarly,  $|a^t(J^i)^{-1} E_{\theta^i}(Y_j^i)|^h = O(p_i^{3h})$ . Hence, we have

$$\begin{aligned} E_{\theta^i} |a^t V_j^i|^3 &= E_{\theta^i} |a^t(J^i)^{-1} Y_j^i - a^t(J^i)^{-1} E_{\theta^i}(Y_j^i)|^3 \\ &\leq E_{\theta^i} |a^t(J^i)^{-1} Y_j^i|^3 + 3|a^t(J^i)^{-1} E_{\theta^i}(Y_j^i)| E_{\theta^i} (a^t(J^i)^{-1} Y_j^i)^2 \\ &\quad + 3[a^t(J^i)^{-1} E_{\theta^i}(Y_j^i)]^2 E_{\theta^i} |a^t(J^i)^{-1} Y_j^i| + |[a^t(J^i)^{-1} E_{\theta^i}(Y_j^i)]^3| \quad (25) \\ &= O(p_i^9) = O(p^9). \end{aligned}$$

A similar argument deduces  $E_{\theta^i} |a^t V_j^i|^4 = O(p^{12})$ . By the definition  $B_{1n}^i(c)$  and  $B_{2n}^i(c)$ , the desired result follows. ■

**Proof of Lemma 8.7** Let  $q_*^i(\theta^i)$  be the posterior distribution of  $\theta^i$ . Therefore, we have that

$$\begin{aligned} \tilde{\theta}^i &= \int \theta^i \cdot q_*^i(\theta^i) d\theta^i \\ &= \int (\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1} u^i) q_*^i(\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1} u^i) |n^{-1/2}(J^i)^{-1}| du^i \\ &= \int (\theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1} u^i) \pi_*^i(u^i) du^i = \theta_0^i + n^{-\frac{1}{2}}(J^i)^{-1} \int u^i \pi_*^i(u^i) du^i. \end{aligned}$$

It follows  $\sqrt{n} J^i(\tilde{\theta}^i - \theta_0^i) = \int u^i \pi_*^i(u^i) du^i$ . On the other hand, the following equations hold

$$\int u^i \phi(u^i; \Delta_n^i, I_{S_i}) du = \int (u^i - \Delta_n^i) \phi(u^i; \Delta_n^i, I_{S_i}) du^i + \Delta_n^i \int \phi(u^i; \Delta_n^i, I_{S_i}) du^i = \Delta_n^i.$$

We thus have

$$\sqrt{n} J^i(\tilde{\theta}^i - \theta_0^i) - \Delta_n^i = \int u^i [\pi_*^i(u^i) - \phi(u^i; \Delta_n^i, I_{S_i})] du^i.$$

■

**Proof of Lemma 8.8** Since  $\|\gamma^i\| < \log p$  and  $\frac{\log p}{\sqrt{n}} \rightarrow 0$  by Condition (4\*), then  $\|\frac{\gamma^i}{\sqrt{n}}\| \leq \eta$  where  $\eta$  as given in Lemma 8.1 is the size of the neighborhood for  $\gamma^i$ .

Therefore, by Lemma 8.1, there exists a constant  $C_2$  such that  $\left| \frac{\partial^3 G_{\bar{U}_j^i}^i(\frac{\gamma^i}{\sqrt{n}})}{\partial \gamma_k^i \partial \gamma_l^i \partial \gamma_m^i} \right| \leq C_2$ . We also have

$$\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i = O\left(\frac{\log p}{\sqrt{n}}\right).$$

According to Condition (4\*),  $\frac{\log p}{\sqrt{n}} = o(1)$ . Therefore, for any arbitrary constant  $a$  such that  $a^2 > 1$ ,  $\frac{1}{3} \frac{C_2}{\sqrt{n}} \sum_{m=1}^{S_i} \gamma_m^i \leq a^2 - 1$ . Following the argument similar to that of Lemma 8.2, we obtain

$$\log E[e^{(\gamma^i)^t \eta^i}] \leq a^2 \|\gamma^i\|^2 / 2.$$

■

**Proof of Lemma 8.9** According to Lemma 8.8, we have

$$\log(E\{\exp[(\gamma^i)^t \Delta_n^i]\}) \leq a^2 \|\gamma^i\|^2 / 2 \quad \text{for} \quad \|\gamma^i\| \leq \log p$$

where  $a$  is a constant with  $a^2 > 1$ . Condition (5) implies  $a^2 \log^2 p > S_i$ . Let  $x_c^i$  be defined as in the proof of Lemma 8.3. Then  $x_c^i > \frac{1}{4} a^2 \log^2 p$  and let  $x = \frac{1}{6} \log^2 p$ , then we have  $\frac{S_i}{6.6} < x < x_c^i$ . Following similar argument as in the proof of Lemma 8.3, we can obtain that  $P(\|\Delta_n^i\|^2 \geq 3a^2 \log^2 p) \leq 10.4e^{-\frac{1}{6} \log^2 p}$ . ■

**Proof of Lemma 8.10** Let  $B^i(\theta^i) = \psi'(\theta^i) - \bar{Y}^i$  be the negative of the score function. Then the MLE  $\hat{\theta}^i$  satisfy the likelihood equation  $B^i(\hat{\theta}^i) = 0$ . Let  $b_n = \frac{\sqrt{3}a \log p}{\sqrt{n}} \frac{\lambda_{\max}^{\frac{1}{2}}(F^i)}{\lambda_{\min}(F^i)}$  with  $a^2 > 1$ . We are going to show with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ , for any  $\theta^i$  on the ball  $\|\theta^i - \theta_0^i\| = 1.2b_n$ , we have

$$(\theta^i - \theta_0^i)^t B^i(\theta^i) > 0. \quad (26)$$

Because that according to Theorem 6.3.4 of Ortega & Rheinboldt [1970], this will imply that there exists a root of  $B^i(\hat{\theta}^i) = 0$  inside the ball  $\|\theta^i - \theta_0^i\| \leq 1.2b_n$  and thus with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ ,  $\|J^i(\hat{\theta}^i - \theta_0^i)\| \leq \lambda_{\max}^{\frac{1}{2}}(F^i) 1.2b_n \leq c' \frac{\log p}{\sqrt{n}}$ . To complete the proof, it now suffices to show the inequality (26) holds. Based on (2.3) in Proposition 2.1 of Portnoy [1988], we have

$$\begin{aligned} (\theta^i - \theta_0^i)^t B^i(\theta^i) &= (\theta^i - \theta_0^i)^t (\psi'(\theta^i) - \bar{Y}^i) \\ &= (\theta^i - \theta_0^i)^t \mu^i + (\theta^i - \theta_0^i)^t \psi''(\theta_0^i) (\theta^i - \theta_0^i) \\ &\quad + \frac{1}{2} E_{\bar{\theta}^i} [(\theta^i - \theta_0^i)^t V_j^i]^3 - (\theta^i - \theta_0^i)^t \bar{Y}^i \\ &= -(\theta^i - \theta_0^i)^t [\bar{Y}^i - \mu^i] + (\theta^i - \theta_0^i)^t \psi''(\theta_0^i) (\theta^i - \theta_0^i) \\ &\quad + \frac{1}{2} E_{\bar{\theta}^i} [(\theta^i - \theta_0^i)^t V_j^i]^3 \\ &= \text{term1} + \text{term2} + \text{term3}, \end{aligned}$$

where  $\mu^i = \psi'(\theta_0^i)$ ,  $V_j^i$  is defined as in (7) of Section 4 and  $\bar{\theta}^i$  is a point on the line segment between  $\theta^i$  and  $\theta_0^i$ . It is easy to see that

$$term2 \geq \lambda_{min}(F^i) \|(\theta^i - \theta_0^i)\|^2.$$

For *term3*, under Condition (5), from (24), (25) and Lemma 9.1, we see that

$$\sup\{|E_\theta(a^t V_j^i)^3| : \|a\| = 1, \|\theta^i - \theta_0^i\| = 1.2b_n\} \leq \sup\{|E_\theta(a^t V_j^i)^3| : \|a\| = 1, \|\theta^i - \theta_0^i\| \leq 1.2b_n\}$$

is bounded. Since  $b_n \rightarrow 0$ , then  $\frac{0.1}{b_n} \lambda_{min}(F^i) \rightarrow \infty$ . Therefore,

$$\sup\{|E_\theta(a^t V_j^i)^3| : \|a\| = 1, \|\theta^i - \theta_0^i\| = 1.2b_n\} \leq \frac{0.1}{b_n} \lambda_{min}(F^i).$$

It follows

$$term3 \geq -\frac{0.05}{b_n} \lambda_{min}(F^i).$$

In *term1*, there is a random term  $\bar{Y}^i - \psi'(\theta_0^i)$ . We will now show that

$$term1 \geq \frac{\sqrt{3a} \log p}{\sqrt{n}} \lambda_{max}^{\frac{1}{2}}(F^i) \|\theta^i - \theta_0^i\|$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . According to Lemma 8.9, we have  $\|\Delta_n^i\|^2 \leq 3a^2 \log^2 p$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . Furthermore, since

$$\|\Delta_n^i\|^2 = \|\sqrt{n}(J^i)^{-1}(\bar{Y}^i - \mu^i)\|^2 = n(\bar{Y}^i - \mu^i)^t (F^i)^{-1}(\bar{Y}^i - \mu^i) \geq \frac{n}{\lambda_{max}(F^i)} \|\bar{Y}^i - \mu^i\|^2,$$

then  $\frac{n}{\lambda_{max}(F^i)} \|\bar{Y}^i - \mu^i\|^2 \leq 3a^2 \log^2 p$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ .

It implies  $\|\bar{Y}^i - \mu^i\|^2 \leq \frac{\sqrt{3a} \log p}{\sqrt{n}} \lambda_{max}^{\frac{1}{2}}(F^i)$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ .

Consequently,  $term1 \geq \frac{\sqrt{3a} \log p}{\sqrt{n}} \lambda_{max}^{\frac{1}{2}}(F^i) \|\theta^i - \theta_0^i\|$ . Combining the above results, on the ball of  $\|\theta^i - \theta_0^i\| = 1.2b_n$ , we have

$$\begin{aligned} & (\theta^i - \theta_0^i)^t B^i(\theta^i) \\ & \geq -\frac{\sqrt{3a} \log p}{\sqrt{n}} \lambda_{max}^{\frac{1}{2}}(F^i) \|\theta^i - \theta_0^i\| + \lambda_{min}(F^i) \|(\theta^i - \theta_0^i)\|^2 - \frac{0.05}{b_n} \lambda_{min}(F^i) \|\theta^i - \theta_0^i\|^3 \\ & = -3a^2 \sqrt{\frac{\log^2 p}{n}} \lambda_{max}^{\frac{1}{2}}(F^i) 1.2b_n + \lambda_{min}(F^i) (1.2b_n)^2 - \lambda_{min}(F^i) \frac{0.05}{b_n} (1.2b_n)^3 \\ & \geq \lambda_{min}(F^i) b_n^2 [-1.2 + (1.2)^2 - 0.05(1.2)^3] > 0 \end{aligned}$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . Therefore, we proved that  $\|\hat{\theta}^i - \theta_0^i\| \leq 1.2b_n$  with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . It follows

$$\|J^i(\hat{\theta}^i - \theta_0^i)\| \leq \lambda_{max}^{\frac{1}{2}}(F^i) \|\hat{\theta}^i - \theta_0^i\| \leq \lambda_{max}^{\frac{1}{2}}(F^i) 1.2b_n$$

with probability greater than  $1 - 10.4 \exp\{-\frac{1}{6} \log^2 p\}$ . ■

**Lemma 9.1** Let  $\theta_j^i$  be the  $j$ -th element of  $\theta^i$ ,  $i = 1, 2, \dots, p$ . Under Condition (2), for  $\|\theta^i - \theta_0^i\| \leq \varepsilon_1$ , we have that  $|\theta_j^i| \leq \varepsilon_1 + \kappa_2$ .

**Proof** Let  $\theta_{j,0}^i$  be the  $j$ -th element of  $\theta_0^i$ ,  $i = 1, 2, \dots, p$ . Since  $\|\theta^i - \theta_0^i\| \leq \varepsilon_1$ , then  $\sqrt{\sum_{j=1}^{S_i} (\theta_j^i - \theta_{j,0}^i)^2} \leq \varepsilon_1$ . Therefore, for any  $j \in \{1, 2, \dots, S_i\}$ , we have

$$\sqrt{(\theta_j^i - \theta_{j,0}^i)^2} \leq \sqrt{\sum_{j=1}^{S_i} (\theta_j^i - \theta_{j,0}^i)^2} \leq \varepsilon_1.$$

It implies  $|\theta_j^i - \theta_{j,0}^i| \leq \varepsilon_1$ . By Proposition 9.2, under Condition (2), we have  $|\theta_{j,0}^i| \leq \kappa_2$ . It follows  $|\theta_j^i| \leq \varepsilon_1 + \kappa_2$ . ■

**Lemma 9.2** Let  $Y_j^i$  be defined in (5) of Section 4 and denote  $Y_j^i = (Y_{j1}, Y_{j2}, \dots, Y_{jS_i})^t$ , under Condition (2) and  $\|\theta^i - \theta_0^i\| \leq \varepsilon_1$ , we have  $E_{\theta^i} [|Y_{jk_1}^i| \cdots |Y_{jk_h}^i|]$  is bounded for  $h = 1, 2, 3, 4$ .

**Proof** According to Lemma 9.1, each element of  $\theta^i$  is bounded. Since  $Y_{jk}^i = -\frac{1}{2} \text{tr}(\delta_k^i X_j^i (X_j^i)^t)$ , by Isserlis' Theorem, the moments of every entry of  $X_j^i (X_j^i)^t$  is finite. By Condition (3),  $E_{\theta^i}(Y_{jk}^i)$  is bounded and  $E_{\theta^i}(Y_{jk}^i)^2$  is also bounded. By Hölder's inequality, we have  $E_{\theta^i} [|XY|] \leq (E_{\theta^i} [|X|^p])^{\frac{1}{p}} (E_{\theta^i} [|Y|^q])^{\frac{1}{q}}$ . Therefore, when  $h = 1$ ,  $E_{\theta^i} (|Y_{jk_1}^i|) \leq [E_{\theta^i} (Y_{jk_1}^i)^2]^{\frac{1}{2}}$  is bounded. When  $h = 2$ , we have

$$E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i|) \leq [E_{\theta^i} (Y_{jk_1}^i)^2]^{\frac{1}{2}} [E_{\theta^i} (Y_{jk_2}^i)^2]^{\frac{1}{2}}.$$

It follows  $E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i|)$  is bounded. When  $h = 3$ , we have

$$E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i| |Y_{jk_3}^i|) \leq [E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i|)^2]^{\frac{1}{2}} [E_{\theta^i} (Y_{jk_3}^i)^2]^{\frac{1}{2}}.$$

Since  $E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i|)$  is bounded, then  $E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i|)^2$  is also bounded. Therefore,  $E_{\theta^i} (|Y_{jk_1}^i| |Y_{jk_2}^i| |Y_{jk_3}^i|)$  is bounded. Consequently,  $E_{\theta^i} [|Y_{jk_1}^i| \cdots |Y_{jk_h}^i|]$  is bounded for  $h = 1, 2, 3, 4$ . ■

**Proposition 9.7** Let  $E$  be a Euclidean space and let  $F \subset E$  be a linear subspace. Let  $p_F$  denote the orthogonal projection of  $E$  onto  $F$ . Let  $g$  be a linear symmetric operator  $g : E \rightarrow E$  and consider the linear application  $f$  of  $F$  into itself defined by

$$f : x \in F \rightarrow f(x) = p_F \circ g(x)$$

Then, we have that if  $\mu_1 < \mu_2 < \cdots < \mu_m$  are the eigenvalues of  $g$  and  $\lambda_1 < \lambda_2 < \cdots < \lambda_n$  are the eigenvalues of  $f$ ,  $n < m$ , then for any  $j = 1, 2, \dots, n$ , the following

inequalities hold

$$\mu_1 \leq \lambda_j \leq \mu_m.$$

**Proof** We prove is first for  $m = \dim(F) = \dim(E) - 1$ . Let  $e = (e_1, e_2, \dots, e_n)$  be an orthonormal basis of  $F$  such that basis the matrix representative of  $f$  is a diagonal  $[f]_e^e = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  and let  $e_0 \in E$  be such that  $e' = (e_0, e_1, e_2, \dots, e_n)$  is an orthonormal basis of  $E$ . Then in that basis, the matrix representative of  $g$  is

$$[g]_{e'}^{e'} = \begin{pmatrix} a & b_1 & \cdots & b_n \\ b_1 & \lambda_1 & 0 & 0 \\ \cdots & \cdots & \ddots & 0 \\ b_n & \cdots & 0 & \lambda_n \end{pmatrix}.$$

We see here that the matrix representative of  $f$  is a submatrix of the matrix representative of  $g$ . By the interlacing property of the eigenvalues, we have

$$\mu_1 \leq \lambda_j \leq \mu_{n+1} \quad j = 1, 2, \dots, n.$$

If  $\dim(E) - \dim(F) > 1$ , we iterate the process by induction on  $\dim(E) - \dim(F)$  and complete the proof. ■

**Lemma 9.3** For any  $i \in \{1, 2, \dots, p\}$ , let  $T^i(\gamma^i)$  be a symmetric matrix with dimension  $p_i$ . Then there exists a constant  $\eta$  such that with  $\|T^i(\gamma^i)\|_F \leq \eta$ , the matrix  $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$  is positive definite.

**Proof** If we want to show  $I_{p_i} + T^i(\gamma^i)\Sigma_0^i$  is positive definite, it is equivalent to show for any non zero vector  $z$  with dimension  $p_i$ ,  $z^t(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)z$  is positive. By Cauchy-Schwarz inequality, we have

$$\begin{aligned} | \langle T^i(\gamma^i)z, \Sigma_0^i z \rangle | &\leq \|T^i(\gamma^i)z\| \times \|\Sigma_0^i z\| \leq \|T^i(\gamma^i)\| \times \|z\| \times \|\Sigma_0^i\| \times \|z\| \\ &\leq \|z\|^2 \|T^i(\gamma^i)\|_F \times \frac{1}{\kappa_1} \leq \eta \|z\|^2 \frac{1}{\kappa_1}. \end{aligned}$$

Therefore,

$$\begin{aligned} z^t[I_{p_i} + T^i(\gamma^i)\Sigma_0^i]z &= z^t I_{p_i} z + z^t T^i(\gamma^i)\Sigma_0^i z = \|z\|^2 + \langle T^i(\gamma^i)z, \Sigma_0^i z \rangle \\ &\geq \|z\|^2 - \|z\|^2 \eta \frac{1}{\kappa_1} = (1 - \eta \frac{1}{\kappa_1}) \|z\|^2. \end{aligned}$$

We can thus choose a constant  $\eta$ , such that  $\eta < \kappa_1$ . It follows  $z^t(I_{p_i} + T^i(\gamma^i)\Sigma_0^i)z \geq \|z\|^2 > 0$  when  $\|T^i(\gamma^i)\|_F \leq \eta$ . ■



**Lemma 9.4** *Let  $K = (K_{ij})_{1 \leq i, j \leq p}$  be  $p \times p$  positive semi-definite matrix. Then*

$$\|K\|_F \leq \text{tr}^2(K).$$

**Proof** Because  $K$  is positive semi-definite, we have  $K_{ij}^2 \leq K_{ii}K_{jj}$ . Thus

$$\sum_{1 \leq i, j \leq p} K_{ij}^2 \leq \sum_{1 \leq i, j \leq p} K_{ii}K_{jj} = \left( \sum_{1 \leq i \leq p} K_{ii} \right) \left( \sum_{1 \leq j \leq p} K_{jj} \right) = \left( \sum_{1 \leq i \leq p} K_{ii} \right)^2.$$

■

In order to prove the following Theorem, we start from a finite dimensional real linear space  $E$  of dimension  $n$  (thus isomorphic to  $\mathbb{R}^n$  but we prefer to avoid the use of artificial coordinates). We denote by  $E^*$  its dual space, that means the set of linear applications  $\theta : E \mapsto \mathbb{R}$ . We denote  $\langle \theta, x \rangle = \theta(x)$ . If  $E$  is Euclidean, the dual  $E^*$  is identified with  $E$  and  $\langle \theta, x \rangle$  is the scalar product.

Consider a non empty open convex cone  $C$  with closure  $\overline{C}$  such that  $C$  is proper, that is to say such that

$$\overline{C} \cap (-\overline{C}) = \{0\}.$$

The dual cone of  $C$  is

$$C^* = \{\theta \in E^* ; \langle \theta, x \rangle \geq 0 \ \forall x \in \overline{C} \setminus \{0\}\}.$$

This is a standard result of convex analysis that  $C^*$  is not empty (Faraut & Korányi [1994]). In general the description of  $C^*$  is a non trivial matter.

A polynomial  $P$  on  $E$  is a function  $P : E \mapsto \mathbb{R}$  such that if  $e = (e_1, \dots, e_n)$  is a basis of  $E$  and if  $x = x_1 e_1 + \dots + x_n e_n \in E$  then  $P(x)$  is a polynomial with respect to the real variables  $(x_1, \dots, x_n)$ . Needless to say the definition does not depend on the particular chosen basis  $e$ . A polynomial  $P$  is homogeneous of degree  $k$  if for all  $\lambda \in \mathbb{R}$  and all  $x \in E$  we have

$$P(\lambda x) = \lambda^k P(x).$$

**Theorem 9.1** *Let  $C$  be an open convex and proper cone of  $E$ , let  $P$  be a homogeneous polynomial on  $E$  of degree  $k$  and let  $\alpha > -n/k$ . We assume that  $P(x) > 0$  on  $C$ . We choose a Lebesgue measure  $dx$  on  $E$ . For  $\theta \in E^*$  consider the integral*

$$L(\theta) = \int_C e^{-\langle \theta, x \rangle} P(x)^\alpha dx \leq \infty.$$

*If  $\theta \notin C^*$  the integral  $L(\theta)$  diverges. If  $\theta \in C^*$  denote  $H_1 = \{x \in E ; \langle \theta, x \rangle = 1\}$ . Then  $\overline{C} \cap H_1$  is compact. In this case  $\theta \in C^*$ , the integral  $L(\theta)$  is finite if and only if*

$\int_{C \cap H_1} P(x)^\alpha dx$  is finite. Furthermore

$$L(\theta) = \Gamma(\alpha k + n) \int_{C \cap H_1} P(x)^\alpha dx. \quad (27)$$

**Proof (personal communication from G. Letac)** Suppose that  $\theta_0 \in C^*$  and let us show (27). Consider the affine hyperplanes  $H_1$  and  $H_0$  of  $E$  defined by

$$H_1 = \{x \in E ; \langle \theta_0, x \rangle = 1\}, \quad H_0 = \{x \in E ; \langle \theta_0, x \rangle = 0\}.$$

The convex set  $\overline{C} \cap H_1$  is compact. To see this let us choose an arbitrary scalar product on  $E$ . Observe that the function  $u \mapsto \langle \theta_0, u \rangle$  defined on the intersection of  $\overline{C}$  with the unit sphere of  $E$  is continuous and reaches a minimum  $m > 0$  since the set of definition is compact. Thus for all  $x \in \overline{C} \cap H_1$  we have

$$\|x\| \leq \frac{1}{m} \langle \theta_0, x \rangle = \frac{1}{m}$$

and the closed set  $\overline{C} \cap H_1$  is also bounded, thus compact.

We fix now  $h_1 \in H_1$  and we write any element  $x$  of  $E$  in a unique way as  $x = x_0 + x_1 h_1$  where  $x_1$  is a number and  $x_0$  is in  $H_0$ . If  $E$  is Euclidean, a natural choice for  $h_1$  is  $\theta_0 / \|\theta_0\|^2$  although other choices would be possible. We also write  $x = (x_0, x_1)$  for short. We denote by  $K \subset H_0$  the set of  $x_0$  such that  $x_0 + h_1 = (x_0, 1)$  is in  $\overline{C} \cap H_1$ . Note that  $K$  is also compact. We get that  $x = (x_0, x_1)$  is in  $\overline{C} \setminus \{0\}$  if and only if  $y = x_0/x_1 \in K$  and  $x_1 > 0$ . To see this denote

$$C_1 = \{(x_0, x_1) ; y = x_0/x_1 \in K, x_1 > 0\}.$$

The inclusion  $C_1 \subset \overline{C} \setminus \{0\}$  is obvious as well as  $\overline{C} \setminus H_0 \subset C_1$ . However if  $(x_0, 0)$  is in  $\overline{C} \cap H_0$  and if  $x_0 \neq 0$  this implies that  $(\lambda x_0, 0)$  is in  $\overline{C} \cap H_0$  for all  $\lambda > 0$  and thus  $\lambda x_0 \in K$  for all  $\lambda > 0$ : this contradicts the compactness of  $K$ . As a result  $(x_0, 0)$  in  $\overline{C} \cap H_0$  implies  $x_0 = 0$ . This implies  $\overline{C} \setminus \{0\} = \overline{C} \setminus H_0$  and thus  $\overline{C} \setminus \{0\} = C_1$

We are now in position to make the change of variable  $(x_0, x_1) \mapsto (y = x_0/x_1, x_1)$  in the integral  $L(\theta_0)$  with an easy Jacobian, since  $\dim H_0 = n - 1$ :

$$dx = dx_0 dx_1 = x_1^{n-1} dy dx_1.$$

We get

$$L(\theta_0) = \int_C e^{-x_1} P(x_0, x_1)^\alpha dx_0 dx_1 = \int_K I(y) dy$$

where

$$I(y) = \int_0^\infty e^{-x_1} P(yx_1, x_1)^\alpha x_1^{n-1} dx_1 = P(y, 1)^\alpha \Gamma(\alpha k + n) = P(y + h_1)^\alpha \Gamma(\alpha k + n)$$

from the homogeneity of the polynomial  $P$ . Thus (27) is proved.

Suppose that  $\theta_0 \notin C^*$ . This is saying that there exists  $x_0 \in C$  such that  $\langle \theta_0, x_0 \rangle \leq 0$ . Let us show that  $L(\theta_0) = \infty$ . Since  $C$  is open we may assume that  $\langle \theta_0, x_0 \rangle < 0$ . Choose an arbitrary scalar product on  $E$ . There exists  $\epsilon$  such that for all  $x$  in  $B = \{x ; \|x - x_0\| < \epsilon\}$  we have  $x \in C$  and  $\langle \theta_0, x \rangle < 0$ . Consider the open subcone  $C_1 = \{\lambda x ; x \in B, \lambda > 0\}$  of  $C$ . We can write

$$L(\theta_0) \geq \int_{C_1} e^{-\langle \theta_0, x \rangle} P(x)^\alpha dx \geq \int_{C_1} P(x)^\alpha dx.$$

Clearly the last integral diverges for  $\alpha \geq 0$ . For  $-n/k < \alpha < 0$  we use the same trick: we parameterize  $C_1$  with the help of the compact set  $\overline{C_1} \cap H_1$  by considering the compact set  $K_1$  of  $y \in H_0$  such that  $y + h_1 \in \overline{C_1} \cap H_1$  and we write

$$\int_{C_1} P(x)^\alpha dx = \int_{K_1} \int_0^\infty P^\alpha(x_0, x_1) dx = \int_{K_1} P(y, 1)^\alpha \left( \int_0^\infty x_1^{\alpha k + n - 1} dx_1 \right) dy = \infty.$$

This proves  $\Rightarrow$ . ■

**Lemma 9.5** *For any  $i \in \{1, 2, \dots, p\}$ , there exists a constant  $M_7$  such that*

$$\log \int_{\|\sqrt{n}J^i(\theta^i - \theta_0^i)\|^2 > cM(p)} \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \leq \exp[M_7 p^2 \log p].$$

**Proof** Without loss of generality, let  $\theta_k^i, k = 1, \dots, s_i$ , be the entries of  $K^i$  on the diagonal and  $\theta_k^i, k = s_i + 1, \dots, S_i$ , be the off-diagonal entries. We assume that  $D^i = I_{p_i}$ , which we need later on anyway. Then

$$\begin{aligned} \text{tr}(K^i D^i) &= \text{tr}(K^i I_{p_i}) = \sum_{k=1}^{s_i} \tau_k^i \theta_k^i = \text{tr}(K^i), \\ \pi_0^i(\theta^i; \delta^i, I_{p_i}) &= \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\}, \quad \text{and} \\ \|\theta^i\|^2 &= \sum_{k=1}^{s_i} (\theta_k^i)^2 + \sum_{k=s_i+1}^{S_i} (\theta_k^i)^2 \leq \sum_{k=1}^{S_i} \tau_k^i (\theta_k^i)^2 = \|K^i(\theta^i)\|_F \leq \left(\sum_{k=1}^{s_i} \tau_k^i \theta_k^i\right)^2 \text{ by lemma 9.4,} \end{aligned}$$

where  $\tau_k^i = |v_k^i|$  is the number of elements in the colour class  $v_k^i$ . We therefore have  $\|\theta^i\| \leq \sum_{k=1}^{s_i} \tau_k^i \theta_k^i$ . Let  $H^i$  denote the convex cone  $P_{\mathcal{G}^i}$  for short.

$$\begin{aligned}
& \int_{H^i} \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \\
&= \int_{H^i} \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i \\
&\leq \sqrt{\|F^i\|} \sqrt{n} \int_{H^i} [\|\theta^i\| + \|\theta_0^i\|] \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i \\
&= \sqrt{\|F^i\|} \sqrt{n} \|\theta_0^i\| \int_{H^i} \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i \quad (28)
\end{aligned}$$

$$+ \sqrt{\|F^i\|} \sqrt{n} \int_{H^i} \|\theta^i\| \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i. \quad (29)$$

By Proposition 9.2, we have  $\|\theta_0^i\|^2 \leq S_i \kappa_2^2$ . Furthermore, according to Proposition 9.1, we have  $\|F^i\| \leq \frac{1}{\kappa_1^2}$ . We therefore need to find upper bounds for the integrals in (28) and (29). These two integrals are of the type  $\int_{H^i} f(\theta^i) \alpha^i e^{-tr(\theta^i D^i)} d\theta^i$  where  $f(\theta^i)$  is a homogeneous function of order  $k^i$ . If  $n^i$  is the dimension of the space in which  $H^i$  sits, we use the result of Theorem 9.1. Let  $\bar{D}^i$  be the  $s_i$ -dimensional vector with entries  $\frac{\tau_k^i}{2}$ . We have

$$\begin{aligned}
\int_{H^i} \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i &= \int_{H^i} e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i, \\
\int_{H^i} \|\theta^i\| \exp\left\{-\frac{1}{2} \sum_{k=1}^{s_i} \theta_k^i \tau_k^i + \frac{\delta^i - 2}{2} \log |K^i(\theta^i)|\right\} d\theta^i &\leq \int_{H^i} \left(\sum_{k=1}^{s_i} \tau_k^i \theta_k^i\right) e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i
\end{aligned}$$

and therefore since  $K^i(\theta^i)$  is homogeneous of order  $p_i$ ,  $\sum_{k=1}^{s_i} \tau_k^i \theta_k^i$  is homogeneous of order 1 and 1 is homogeneous of order 0, we have, for  $\alpha^i = \frac{\delta^i - 2}{2}$

$$\begin{aligned}
& \int_{H^i} e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i = \Gamma(\alpha^i p_i + S_i) \int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i, \\
& \int_{H^i} \left(\sum_{k=1}^{s_i} \tau_k^i \theta_k^i\right) e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i \\
&= \Gamma(\alpha^i p_i + 1 + S_i) \int_{H^i \cap H_1^i} \left(\sum_{k=1}^{s_i} \tau_k^i \theta_k^i\right) |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i. \quad (30)
\end{aligned}$$

However, we do not know how to compute the integrals  $\int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i$  and  $\int_{H^i \cap H_1^i} \left(\sum_{k=1}^{s_i} \tau_k^i \theta_k^i\right) |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i$ . The set  $H_1^i = \{\theta^i \mid tr(\bar{D}^i \theta^i) = 1\}$  is  $H_1^i = \{\theta^i \mid$

$\sum_{k=1}^{s_i} \tau_k^i \theta_k^i = 2\}$ . So, we only have one integral,  $\int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i-2}{2}} d\theta^i$ , to compute. But

$$\sum_{k=1}^{s_i} \tau_k^i \theta_k^i = \text{tr}(K^i(\theta^i) I_{p_i}) = \text{tr}(K^i(\theta^i)) = \sum_{j=1}^{p_i} \lambda_j$$

where the  $\lambda_j$  are the eigenvalues of  $K^i(\theta^i)$ . Following the inequality between the arithmetic mean and the geometric mean, on  $H_1^i \cap H^i$ , we have

$$|K^i(\theta^i)| = \left( \prod_i \lambda_j \right) \leq \frac{(\sum_{j=1}^{p_i} \lambda_j)^{p_i}}{p_i^{p_i}} = \frac{2^{p_i}}{p_i^{p_i}}$$

and thus

$$\int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i-2}{2}} d\theta^i \leq \frac{2^{p_i}}{p_i^{p_i}} \int_{H^i \cap H_1^i} d\theta^i. \quad (31)$$

We are now going to use Theorem 9.1 in the reverse direction with  $f(\theta^i) = 1$  in order to evaluate  $\int_{H^i \cap H_1^i} d\theta^i$ . We have

$$\int_{H^i} e^{-\text{tr}(\bar{D}^i \theta^i)} d\theta^i = \Gamma(0 + S_i) \int_{H^i \cap H_1^i} d\theta^i \quad (32)$$

and we are going to majorize  $\int_{H^i} e^{-\text{tr}(\bar{D}^i \theta^i)} d\theta^i$ . We now use the fact that the matrices in  $H^i$  are positive definite, thus we have that, for  $l = s_i + 1, \dots, S_i$ ,  $(\theta_l^i)^2 \leq \theta_{t_l}^i \theta_{u_l}^i$  whenever  $\theta_l^i = K_{jk}^i$ ,  $j \neq k$  and  $\theta_{t_l}^i = K_{jj}^i, \theta_{u_l}^i = K_{kk}^i$ . Since the cone  $H^i$  is included in the cone  $P^i$  of positive definite matrices, we have that, for  $l = s_i + 1, \dots, S_i$ ,  $(\theta_l^i)^2 \leq \theta_{t_l}^i \theta_{u_l}^i$  whenever  $\theta_l^i = K_{jk}^i$ ,  $j \neq k$  and  $\theta_{t_l}^i = K_{jj}^i, \theta_{u_l}^i = K_{kk}^i$  and thus we can write

$$\begin{aligned} \int_{H^i} e^{-\text{tr}(\bar{D}^i \theta^i)} d\theta^i &\leq \int_0^{+\infty} \dots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}_r^i \theta_r^i} \left[ \prod_{l=s_i+1}^{S_i} \int_{-\sqrt{\theta_{t_l}^i \theta_{u_l}^i}}^{\sqrt{\theta_{t_l}^i \theta_{u_l}^i}} d\theta_l^i \right] \prod_{r=1}^{s_i} d\theta_r^i \\ &= \int_0^{+\infty} \dots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}_r^i \theta_r^i} \left[ \prod_{l=s_i+1}^{S_i} 2\sqrt{\theta_{t_l}^i \theta_{u_l}^i} \right] \prod_{r=1}^{s_i} d\theta_r^i. \end{aligned}$$

Since we have assumed that  $D^i$  is equal to the identity,  $\bar{D}_r^i = \frac{\tau_r^i}{2}$  with the  $\tau_r^i$  being bounded. Then

$$\begin{aligned} &\int_0^{+\infty} \dots \int_0^{+\infty} e^{-\sum_{r=1}^{s_i} \bar{D}_r^i \theta_r^i} \prod_{r=1}^{s_i} \left[ \prod_{l=s_i+1}^{S_i} 2\sqrt{\theta_{t_l}^i \theta_{u_l}^i} \right] d\theta_r^i \\ &= 2^{S_i-s_i} \prod_{r=1}^{s_i} \int_0^{+\infty} (\theta_r^i)^{\frac{k_r^i}{2}} e^{-\frac{\tau_r^i \theta_r^i}{2}} d\theta_r^i = 2^{S_i-s_i} \prod_{r=1}^{s_i} \left( \frac{2}{\tau_r^i} \right)^{\frac{k_r^i}{2}+1} \Gamma\left(\frac{k_r^i}{2} + 1\right) \end{aligned}$$

where  $k_r^i$  is the number of  $t_l$  or  $u_l$  equal to  $r$  in the  $i$ -th local model. From the majorization above, (32), (31) and (30) successively, we obtain the following inequalities

$$\begin{aligned}
\int_{H^i \cap H_1^i} d\theta^i &\leq \frac{1}{\Gamma(S_i)} 2^{S_i - s_i} \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right), \\
\int_{H^i \cap H_1^i} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right), \\
\int_{H^i} \left(\sum_{k=1}^{S_i} \tau_k^i \theta_k^i\right) e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i + 1} \Gamma(\alpha^i p_i + 1 + S_i)}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right), \\
\text{and } \int_{H^i} e^{-tr(\bar{D}^i \theta^i)} |K^i(\theta^i)|^{\frac{\delta^i - 2}{2}} d\theta^i &\leq \frac{2^{S_i - s_i + p_i} \Gamma(\alpha^i p_i + S_i)}{p_i^{p_i} \Gamma(S_i)} \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int \|\sqrt{n}J(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \\
&\leq n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} \left[ \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right) \right] (M_0 p_i \Gamma(\alpha^i p_i + S_i) + 2\Gamma(\alpha^i p_i + 1 + S_i)) \\
&\leq n^{\frac{1}{2}} \frac{1}{\kappa_1} \frac{2^{S_i - s_i + p_i}}{p_i^{p_i} \Gamma(S_i)} \left[ \prod_{r=1}^{s_i} \left(\frac{2}{\tau_r^i}\right)^{\frac{k_r^i}{2} + 1} \Gamma\left(\frac{k_r^i}{2} + 1\right) \right] M_2 p_i \Gamma(\alpha^i p_i + 1 + S_i),
\end{aligned}$$

where  $M_0$  and  $M_2$  are constants. Therefore,

$$\begin{aligned}
&\log \left\{ \int \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \right\} \\
&\leq \frac{1}{2} \log n - \log \kappa_1 + (S_i - s_i + p_i) \log 2 - p_i \log p_i - \log \Gamma(S_i) + \log M_2 + \log p_i \\
&\quad + \log \Gamma(\alpha^i p_i + 1 + S_i) + \sum_{r=1}^{s_i} \left[ \left(\frac{k_r^i}{2} + 1\right) \log \frac{2}{\tau_r^i} + \log \Gamma\left(\frac{k_r^i}{2} + 1\right) \right].
\end{aligned}$$

Since  $\log n$  and  $\log p$  is the same order and  $k_r^i \leq p_i$ , we have that

$$\begin{aligned}
&\exp[\log \left\{ \int \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \right\}] \\
&\leq \exp\left[\frac{1}{2} \log p + (S_i - s_i + p_i) \log 2 + \log p_i + \log \Gamma(\alpha^i p_i + 1 + S_i) + p_i \left(\frac{p_i}{2} + 1\right) \log 2 \right. \\
&\quad \left. + p_i \log \Gamma\left(\frac{p_i}{2} + 1\right) + M_3\right],
\end{aligned}$$

where  $M_3$  is a constant. By Sterling's approximation, we have  $\log n! = n \log n + O(\log n)$ .

Therefore, there exist two constant  $M_5$  and  $M_6$  such that

$$\log \Gamma(\alpha^i p_i + 1 + S_i) \leq \log \Gamma(\alpha^i p_i + 1 + \frac{p_i(p_i + 1)}{2}) \leq \log(2\alpha^i p_i + p_i^2)! \leq M_5 p_i^2 \log p_i$$

and

$$\log \Gamma\left(\frac{p_i}{2} + 1\right) \leq \log p_i! \leq M_6 p_i \log p_i.$$

Combining all results above, we obtain that

$$\begin{aligned} & \exp \left( \log \left\{ \int \|\sqrt{n}J^i(\theta^i - \theta_0^i)\| \pi_0^i(\theta^i) d\theta^i \right\} \right) \\ \leq & \exp \left( \frac{1}{2} \log p + \left[ \frac{p_i(p_i + 1)}{2} + p_i \right] \log 2 + \log p_i + M_5 p_i^2 \log p_i + p_i \left( \frac{p_i}{2} + 1 \right) \log 2 \right. \\ & \left. + M_6 p_i^2 \log p_i + M_3 \right) \leq \exp[M_7 p^2 \log p], \end{aligned}$$

where  $M_7$  is a constant. ■